



AI-Generated Feedback in English Writing: Proficiency Development and Learner Perceptions

Jian Wang^{1*}, Xiaohong Zhang² 

¹School of Foreign Languages and Cultures, Geely University of China, Chengdu, China

²School of Foreign Languages and Cultures, Panzhihua University, Panzhihua, China

E-mail addresses: wangjian@guc.edu.cn

ARTICLE INFO

Article history:

Received June 7, 2025

Revised September 07, 2025

Accepted December 12, 2025

Available online December 17, 2025

Kata Kunci :

Umpan Balik yang dihasilkan AI, Umpan Balik Guru, Instruksi Menuliskan Bahasa Inggris, Mahasiswa Perguruan Tinggi

Keywords:

AI-generated Feedback, Teacher Feedback, English Writing Instruction, College Students

ABSTRAK

Studi ini dilakukan dalam konteks pendidikan bahasa Inggris non-native di Cina. Dibandingkan efektivitas umpan balik AI dan guru dalam meningkatkan kemampuan menulis siswa. Diadakan survei kuisisioner dan wawancara semi-struktural untuk mengecek respon dan sikap siswa terhadap umpan balik AI. Hasil menunjukkan skor post-test kedua kelompok lebih tinggi dibanding pretest. Tidak ada perbedaan signifikan pada kinerja final mereka, menunjukkan efektivitas seragam antara umpan balik AI dan guru. Kelompok eksperimen menunjukkan penerimaan umpan balik AI yang sedang, pendapat berbeda tergantung tingkat keahlian. Semua siswa setuju penggabungan umpan balik guru dan AI lebih bermanfaat untuk pengembangan menuliskan. Studi menyimpulkan bahwa kedua metode umpan balik punya manfaat tersendiri, sebaiknya digabung dalam pengajaran penulisan.

ABSTRACT

This study was situated within the context of college-level English-as-a-foreign-language education in China. It compared the effectiveness of AI-generated feedback and teacher feedback in improving English learners' writing proficiency. A questionnaire survey and semi-structured interviews were also conducted to examine learners' acceptance of and attitudes toward AI feedback. Results revealed that both the experimental group (receiving AI feedback) and the control group (receiving teacher feedback) showed significantly higher post-test scores compared to their pretest scores. Although the mean difference in post-test scores between the two groups exceeded that of the pretest, no statistically significant difference was found in their final performance, suggesting comparable efficacy between AI feedback and teacher feedback in enhancing writing proficiency. Additionally, the experimental group demonstrated moderate acceptance of AI feedback, with perceptions varying across proficiency levels. However, learners universally agreed that combining teacher feedback and AI feedback would yield greater benefits for writing development. The study concludes that both feedback modalities have distinct benefits and misfits, advocating for their integrated—rather than exclusive—use in writing instruction.

1. INTRODUCTION

Writing is a multifaceted endeavor that encompasses multiple processes such as ideation, creativity, social interaction, and psychological cognition (Zhao, 2015). Writing feedback, often referred to as a continuation of the writing process (Wei & Li, 2023), is instrumental in prompting writers to recognize the potential for enhancement in both the structural and substantive aspects of their work. Among various forms of writing feedback, teacher-written feedback stands out as a predominant method. Teachers offer written critiques on students' textual errors to refine their writing skills or elevate the quality of their compositions (Hyland, 2013). However, constraints such as onerous teaching loads and large student cohorts have led to issues with teacher feedback, including its suboptimal quality and lack of specificity (Bai & Hu, 2017). In response to these challenges, automated writing evaluation systems (AWE systems) have come to the fore. These systems, exemplified by the *Pigai* (meaning "correction") platform in China and *Criterion* in the United States, harness natural language processing technologies to pinpoint errors and infelicitous expressions in English learners' writing and offer constructive feedback (Ding & Zou, 2024).

Most recently, with the advent of sophisticated large language models (LLMs), automated writing evaluation is poised to enter a new epoch (Bai & Wei, 2024). Generative artificial intelligence tools or GenAI tools for short, predicated on LLMs like ChatGPT and DeepSeek, are not merely adept at rectifying essay errors but also capable of producing complete essays. With a straightforward prompt, these GenAI tools

can generate essays that adhere to quality benchmarks (Herbold et al., 2023). For English-as-a-foreign-language (EFL) learners, these models can be leveraged to translate essays from their native tongue into English, correct errors en masse, modulate the essay's tone, or incorporate cohesive elements such as discourse markers (Tate et al., 2023). The academic community broadly posits that educational practices must evolve in tandem with GenAI technology's development (Tseng & Warschauer, 2023). Nevertheless, empirical evidence is needed to ascertain whether the integration of GenAI tools into educational practice yields significant pedagogical benefits (Chiu et al., 2023). Therefore, this study investigated the repercussions of AI-generated feedback for EFL learners' writing proficiency and their receptiveness to applying AI feedback in writing instruction.

Effects of AI Feedback on EFL Writing Proficiency

AWE systems, integral to high-stakes English proficiency tests such as the GRE and TOEFL, are designed to batch-score essays. These systems, when juxtaposed with conventional manual scoring methods, demonstrate a marked reduction in subjective bias and variability, thereby enhancing the stability of the scoring accuracy (Grove et al., 2000). In pedagogical contexts, writing instructors may entrust AWE systems with the responsibility of essay grading, thereby freeing up valuable instructional time (Kumar, 2023). Although platforms like *Pigai* and *Criterion* have garnered recognition for their efficacy in evaluating essays, there remains a dearth of scholarly investigation into the efficacy and dependability of GenAI tools in the realm of written text assessment and the willingness of learners to embrace their feedback. AWE systems leverage latent semantic analysis, drawing upon extensive essay corpora to score compositions and provide constructive feedback by statistically gauging the alignment of student essays with established writing norms and semantic contents (Shermis et al., 2013). Conversely, GenAI tools are not confined to domain-specific training but are instead nurtured on a diverse array of texts culled from the internet, raising questions about their suitability as AWE tools that warrant further exploration.

(Ingle & Pack, 2023) have proposed methodologies to improve academic writing proficiency among EFL learners through GenAI tools. They advocate for an interactive approach where learners engage with AI-supported chatbots to refine their writing skills through critical questioning and contemplation of AI-generated content, rather than relying solely on GenAI for topic generation. This human-computer interaction is posited to stimulate creativity and deepen topical comprehension. Additionally, they suggest assigning GenAI specific roles, such as reviewers for conference proposals or instructors in writing workshops, to aid in drafting outlines or providing feedback on draft structures, thus facilitating the writing process. They also recommend utilizing GenAI tools to offer formative feedback on essay coherence, grammar, vocabulary, and tone. AI-supported chatbots can act as knowledgeable interlocutors, tailoring guidance to learners at various stages of the writing journey.

Although some studies have theoretically discussed the potential of GenAI tools in EFL instruction (Chen, 2024), empirical research in writing pedagogy remains limited. Among the extant literature germane to our research subject, (Dai et al., 2023) explored the effects of ChatGPT-provided error correction feedback on undergraduate writing and discovered that AI-provided feedback was not only more comprehensible but also more detailed compared to teacher feedback. They also found that ChatGPT's feedback aligned closely with teacher opinions in certain aspects of writing. (Wei & Li, 2023) conducted a systematic comparison between teacher feedback and ChatGPT-generated writing feedback on a dataset comprising 208 college English essays. Their findings indicated that ChatGPT demonstrated proficiency in several key areas: initial stage tasks such as grammar verification and punctuation correction; intermediate stage tasks including alternative word selection, enhancement of clarity and brevity, and flow improvement; and advanced stage tasks like text summarization. Another study by (Mizumoto & Eguchi, 2023) fed 12,100 essays authored by non-native English speakers into ChatGPT to evaluate its performance in essay assessment and feedback provision. This study revealed that ChatGPT's reliability and accuracy were commendable based on established scoring criteria. (Bai & Wei, 2024) implemented a "composing-comparison-rewriting" three-stage writing instruction model to investigate how EFL learners focus on, integrate, and perceive ChatGPT's reformulations as writing feedback. Their findings indicated that reformulations effectively directed learners to recognize deficiencies in their initial drafts, especially regarding vocabulary selection, and motivated them to incorporate ChatGPT's revision suggestions during the rewriting phase. Furthermore, a direct correlation was identified between the number of examples learners integrated during rewriting and the quantity, type, and quality of feedback they noted during the comparison stage. However, these studies have yet to conduct comparative analyses of the impact of AI feedback on student writing proficiency in contrast with teacher feedback through controlled teaching experiments. (Wang & Zhang, 2025) conducted a longitudinal investigation to assess the impact of incorporating Kimi—a GenAI tool in China—into English writing pedagogy. Their research findings reveal that the participants demonstrated a marked enhancement in their overall writing

proficiency and analytical dimensions. Despite this, there was a notable absence of improvement in the fluency aspect of their writing. Furthermore, the study observed considerable variability in the degree of writing proficiency enhancement among individual learners. However, this study did not undertake a comparative analysis of the efficacy of AI feedback and teacher feedback in augmenting learners' writing skills.

EFL Learners' Perceptions of AI-provided Feedback

The efficacy of technological innovations is often predicated on their adoption and acceptance by end-users (Davis, 1989). A plethora of scholarly inquiry has delved into the perspectives of EFL learners regarding AWE systems. Empirical evidence suggests that both student and instructor perceptions tend to favor traditional teacher feedback over AWE-generated feedback (Huawei & Aryadoust, 2023). However, some investigations have revealed that students frequently encounter greater perplexity when interpreting teacher-provided feedback compared to AWE feedback (Weigle, 2013). (Roscoe et al., 2017) have delineated that the student stance towards AWE feedback is predominantly characterized as "cautiously positive," with this disposition being shaped by a constellation of factors. These include the learners' initial expectations concerning AWE technologies, as well as their firsthand encounters with the feedback mechanisms these systems provide.

In the current academic discourse, the efficacy of AI-generated feedback in writing instruction remains an underexplored domain, with a dearth of research dedicated to this niche. For example, (Dong, 2024) explored EFL learners' perceptions regarding AI feedback and pointed out that learners emphasized the necessity of teacher support. This study underscores the importance of human intervention in writing instruction, advocating for the establishment of clear pedagogical principles, a rigorous assessment of ChatGPT's feedback, and ongoing oversight to bolster the human element in the writing process. (Erisyerico & Fauzan, 2024) explored Indonesian students' perceptions towards AI feedback and teacher feedback. Their study revealed that students generally regarded AI feedback as more efficacious than teacher feedback in the realms of unity, content development, grammar, and mechanics. Conversely, when it came to coherence, the capacity for personalization, and alignment with educational goals, the perceptions of teacher feedback and AI feedback were found to be analogous. (Bai & Wei, 2024) noted that learners generally perceived the feedback from ChatGPT as possessing merit within EFL writing pedagogy, particularly during the essay revision phase. However, they also highlighted the occasional misinterpretation of user intent by ChatGPT. (Amani & Bisriyah, 2025) found that most students acknowledged AI tools as advantageous, particularly in grammatical verification, orthographic correction, and lexical enhancement. Nonetheless, their dependency on AI for conceptualization, thematic determination, and the oversight of writing progression exhibited variability. Some students lauded AI tools for their role in providing feedback, paraphrasing, and translation services; however, others maintained a preference for human mentorship when seeking detailed writing guidance. Some students even voiced concerns that an overreliance on AI could potentially undermine their comprehension and erode their self-assurance in writing.

AI has been confirmed to positively influence the acquisition of foreign languages (Feng, 2025). Nonetheless, the body of research leveraging the Technology Acceptance Model (TAM) to investigate students' acceptance of AI in language learning remains relatively sparse. Initially introduced by (Davis, 1989), TAM posits that the adoption of new technologies is predominantly swayed by users' perceptions of their usefulness and ease of use. Perceived usefulness denotes individuals' acknowledgement of the benefits conferred by technology, which in turn influences their propensity and actions regarding its utilization. Perceived ease of use concerns the perceived simplicity or complexity of employing such technologies. Attitude towards use acts as an intermediary, encapsulating users' comprehensive assessments of technology, while behavioral intention signifies the potential for its adoption. A handful of studies have probed the acceptance of AI in educational settings, with (Liu & Ma, 2023) providing a notable examination of student acceptance of ChatGPT in foreign language learning. Their research indicates that perceived ease of use does not directly affect attitudes toward technology use. Instead, its influence is mediated by perceived usefulness; the more beneficial users perceive the technology to be, the stronger their intention to utilize it, which subsequently serves as a significant predictor of actual use. However, despite these revelations, the acceptance of AI feedback in English writing remains under-scrutinized. In light of this, the present study employed TAM as a framework to explore Chinese university students' perceptions of the application of AI feedback into English writing.

Rationale for The Present Study

This research represents the first quasi-experimental investigation that compares the pedagogical efficacy of Kimi-generated and teacher-provided feedback within an EFL context. While prior studies have

made controlled comparisons between ChatGPT and teacher feedback, these comparisons often fail to reflect the practical constraints of Chinese EFL settings, where access to ChatGPT is severely limited due to technical and institutional barriers. By contrast, Kimi, a Chinese-developed GenAI tool, is widely accessible and user-friendly, making it a more feasible and relevant option in local educational environments.

Furthermore, this study contributes to the broader discourse on AI-generated writing feedback by extending beyond earlier AWE systems that relied on rule-based algorithms (Zheng, 2025). Traditional AWE tools were largely confined to identifying surface-level linguistic errors—such as grammar, spelling, and punctuation—through predefined templates, often lacking the contextual sensitivity required to address higher-order concerns like coherence, argument development, and stylistic appropriateness (Bai & Hu, 2017). In contrast, AI-generated feedback powered by LLMs offers a more nuanced, holistic, and adaptive approach to writing instruction. LLMs are capable of generating contextually relevant suggestions that go beyond error correction. These include proposing alternative phrasings, offering in-depth revisions to enhance clarity and argument strength, customizing feedback based on specific writing genres, and even simulating process-oriented scaffolding that mimics human-like tutoring interactions (Liu et al., 2025). This paradigm shift in automated feedback moves away from a singular focus on error detection and correction, instead embracing a more comprehensive and pedagogically meaningful role in supporting learners' writing development. By addressing the situational and linguistic complexities inherent in authentic writing tasks, AI-generated feedback has the potential to fundamentally reshape the scope and impact of automated writing instruction.

Grounded in the above rationale, this study was guided by the following two research questions:

- a. Can AI feedback enhance EFL learners' English writing proficiency when compared to traditional teacher feedback?
- b. How do EFL learners perceive and accept AI feedback on their writing?

2. METHODS

Research Participants

Participants in this research were drawn from a comprehensive university in southwestern China during the autumn semester of the 2024-2025 academic year. Employing a non-probability sampling technique, we selected a cohort of 75 sophomore English majors from two intact classes, comprising 62 female and 13 male students, aged from 18 to 22 years (Mean = 19.33, $SD = 0.69$). Their English writing proficiency was rated as intermediate based on the June 2024 College English Test Band 4 (CET-4) results, with an average score of 10.67 out of a possible 15 points ($SD = 0.72$), and no statistically significant disparity was observed in CET-4 writing scores between the two classes ($p > 0.05$). Throughout the study, participants were engaged in a 16-week Comprehensive English course designed to improve their competencies in comprehending, synthesizing, and critically evaluating reading materials, as well as articulating personal perspectives on specified topics. A component of their course assessment entailed an integrated reading and writing exercise, necessitating the reading of approximately 200-word texts followed by the composition of at least 200-word essays encompassing a concise summary and a reflective commentary. This task mirrors a requisite segment of the English Majors Band 4 (TEM-4) national examination, a format with which students were already acquainted. Participants were randomly apportioned into two groups: one class functioned as the experimental group (38 individuals, including 7 males and 31 females), while the other served as the control group (37 individuals, with 6 males and 31 females). The control group received feedback from their Comprehensive English instructor, whereas the experimental group was subjected to feedback from the GenAI tool—Kimi.

Research Instruments

This study embarked on a teaching experiment to examine the efficacy of AI feedback in elevating learners' writing proficiency in comparison to traditional teacher feedback. Concurrently, we used a mixed-methods approach to the inquiry into the acceptance and attitudes of learners towards AI feedback. Collectively, the research methodologies included the utilization of the GenAI tool Kimi, two writing tests (one pretest and one post-test), a quantitative questionnaire, and semi-structured interviews to gather comprehensive data.

GenAI tool—Kimi. In this study, the AI assistant Kimi was utilized to provide feedback for the experimental group. Launched by Beijing Moon's Dark Side Technology Co., Ltd. in October 2023, Kimi is capable of processing various inputs, including images and texts, and generating text outputs. To ensure the effectiveness of the feedback, Kimi was programmed to use concise and clear language, assessing students' writing across three dimensions based on TEM-4 grading criteria: content expression ability (i.e., clarity of central argument, development of viewpoints, presentation of arguments, and logical relationships), discourse organization ability (i.e., rationality of overall structure and clarity of paragraph

division), and language proficiency (i.e., accuracy of vocabulary selection, correctness of grammar, and standardization of punctuation usage).

Writing tests. To compare the effectiveness of Kimi's feedback versus teacher feedback, we utilized a pretest (administered on the first day of the course) and a post-test (administered in Week 10) to measure students' language advancement over time. Participants were tasked with composing a 200-word essay within one class period (45 minutes) for both tests whose difficulty level was aligned with the writing section of TEM-4.

Quantitative questionnaire. This study assessed the acceptance of AI feedback among EFL learners in the experimental group by administering a 5-point Likert-type scale (1 = completely disagree, 5 = completely agree). This scale was adapted from the questionnaire of (Li, 2023), which consists of four dimensions: perceived usefulness (5 items), perceived ease of use (5 items), attitude towards use (4 items), and behavioral intention (4 items). We modified the terminology in the questionnaire of (Li, 2023) questionnaire from "AI-based systems" to "AI-generated writing feedback" and from "learning/study" to "English writing." The questionnaire featured items such as "It is easy for me to learn to use AI-generated writing feedback" (perceived ease of use), "Studying with AI-generated writing feedback would improve my English writing efficiency" (perceived usefulness), "Overall, my attitude towards AI-generated writing feedback is positive" (attitude towards use), and "I intend to adopt AI-generated writing feedback for my English writing" (behavioral intention). The overall scale had high reliability ($\alpha = 0.867$) and the reliability coefficients for the four subscales were 0.838, 0.876, 0.893, and 0.849, respectively.

Semi-structured interviews. To discern students' perceptions of Kimi's feedback, a purposive sample of 12 participants from the experimental group was recruited for semi-structured interviews. This qualitative approach complemented the quantitative survey data by capturing nuanced perspectives. The interviewees were stratified by writing proficiency: 4 advanced writers, 4 intermediate writers, and 4 below-average writers, with each subgroup comprising 3 female students and 1 male student. The interview protocol focused on two key themes: (a) participants' evaluations of the quality and utility of Kimi's writing feedback, and (b) their comparative preferences for AI-generated versus human teacher feedback.

Data Collection and Analysis

Initially, a diagnostic assessment was administered to gauge the writing proficiency of students at the commencement of the academic term. Subsequently, over a period of eight weeks, participants were tasked with weekly submissions of a 200-word essay on a prescribed topic. Throughout this period, both groups engaged in Comprehensive English lessons facilitated by a single instructor, with four sessions per week, each session lasting 45 minutes. Kimi was utilized to provide tailored feedback on each essay draft for the experimental group, with responses being dispatched via email within two days to give students ample time for reflection and revision. The control group reviewed and acted upon feedback from their teachers before finalizing their submissions. A post-test was conducted at the end of Week 10.

Both pretest and post-test were rated independently by two seasoned English writing teachers employing holistic scoring criteria derived from the TEM-4 examinations. Cohen's Kappa coefficients indicated substantial agreement between both raters in the pretest ($\kappa = 0.677, p < 0.001$) and post-test ($\kappa = 0.669, p < 0.001$). In week 11, a questionnaire was administered to the experimental class to understand students' acceptance of AI feedback. Following this, a purposive sample of 12 participants was selected for in-depth interviews to capture nuanced perspectives. Interviews were conducted via Tencent Meeting in Mandarin Chinese, each lasting approximately 10 minutes, with recordings made with the consent of the students.

Following data compilation and organization, we implemented a multi-stage analytical protocol: (1) scrutiny for missing data entries and outliers within quantitative datasets; (2) descriptive statistical analysis of the test scores; (3) inferential statistical analyses examining within-group and between-group differences in test outcomes; (4) repeated-measures ANOVA employing a general linear model architecture, considering pretest and post-test timings as within-subjects factors and group allocation (experimental vs. control) as between-subjects factors; (5) preliminary analysis of survey responses regarding learner perceptions of AI feedback; (6) verbatim transcription of interview audios followed by inductive thematic coding. All quantitative analyses were run via SPSS Version 27.0. Interview transcriptions were initially generated through Tencent Meeting's automated transcription feature, followed by the researchers' manual verification.

3. RESULT AND DISCUSSION

Results

Impact of AI-provided Feedback on Writing Proficiency

Upon the validation of quantitative data, it was observed that there were no missing or outlier values. To grasp the general traits of the data set, a descriptive statistical analysis was conducted in this study. As seen in Table 1, the skewness and kurtosis values for all test data lie within the ± 3 range, indicating a normal distribution of the data. This suggests that parametric statistical methods are appropriate for subsequent analyses (Kline, 2023). Additionally, the mean writing scores of the control group in the pretest and post-test were 72.35 ($SD = 5.33$) and 76.69 ($SD = 4.47$), respectively. In contrast, the experimental group's mean scores in both tests were 72.47 ($SD = 4.97$) and 77.48 ($SD = 4.69$), respectively. After the instructional intervention, both groups showed comparable enhancements in English writing proficiency. Notably, the mean score of the control class increased by 4.34 points, whereas that of the experimental group rose by 5.01 points.

Table 1. Descriptive Results of the Pretest and Post-Test

Tests	Groups	Min.	Max.	Mean	SD	Kurtosis	Skewness
Pretest	Control	63	87	72.35	5.33	0.176	0.467
	Experimental	65	85	72.47	4.97	0.581	0.406
Post-test	Control	72	93	76.69	4.47	0.177	0.308
	Experimental	70	91	77.48	4.69	0.721	0.803

As depicted in Table 2, the findings from the mixed 2×2 repeated measures analysis of variance (ANOVA) reveal that the interaction between group and time was statistically non-significant ($F = 2.926$, $p = 0.569$). The effect size analysis indicates that this two-way interaction explained a mere 5.7% of the variance in scores ($\eta^2 = 0.057$). In the between-subjects effect test, no significant difference was observed between the experimental and control groups ($F = 0.339$, $p = 0.232$), suggesting that the type of feedback—whether from a teacher or Kimi—did not substantially impact students' post-test scores. The between-subjects variable (group) only explained 0.7% of the variance ($\eta^2 = 0.007$). Conversely, in the within-subjects effect test, a pronounced difference was identified between pretests and post-tests ($F = 327.327$, $p < 0.001$), with a substantial effect size ($\eta^2 = 0.673$), and the within-subjects variable (time) explained 67.3% of the variance. This finding underscores that there was a marked improvement in writing proficiency over time for students across both groups.

Table 2. Results of Repeated Measures Analysis of Variance

	Type III Sum of Squares	df	Mean Square	F	Sig.	η^2
Between-groups						
Group	0.876	1	0.876	0.339	0.232	0.007
SE	136.542	37	2.736			
With-group						
Time	643.225	1	643.225	327.327	$p < 0.001$	0.673
Time * Group	4.333	1	4.333	2.926	0.569	0.057
SE (Time)	73.827	37	1.441			

Table 3 displays the results of the paired samples t-test, revealing significant disparities in the pre- and post-test scores for both the control and experimental groups ($p < 0.001$), accompanied by moderate effect sizes (Cohen's d values of 0.321 and 0.307, respectively). These results imply that following eight weeks of writing instruction, both groups experienced a notable improvement in their English writing proficiency, aligning with the conclusions drawn from the repeated measures analysis of variance.

Table 3. Results of Paired Samples T-Test

Groups	Mean Difference (pre-post)	t	df	Sig.	Cohen's d
Control	-4.34	-1.673	46	$p < 0.001$	0.321
Experimental	-5.01	-1.724	46	$p < 0.001$	0.307

As presented in Table 4, the independent samples t-test results demonstrate that no statistically significant differences existed in the mean scores between both groups for the pretest and post-test ($p >$

0.05), with minimal effect sizes (Cohen's d values of 0.133 and -0.126, respectively). This suggests that at both assessment points, the writing abilities of students in the two groups were not significantly different. Furthermore, it implies that the distinct methods of writing feedback did not exert a substantial influence on students' proficiency in English writing.

Table 4. Results of Independent Samples T-Test

Tests	Mean Difference (Control-Experimental)	t	df	Sig.	Cohen's d
Pretest	-0.12	-0.752	46	0.072	0.133
Post-test	-0.79	-1.323	46	0.129	-0.126

EFL Learners' Acceptance and Perceptions of AI Feedback

A questionnaire survey was conducted to dig into learners' acceptance of AI feedback. As seen in Table 5, the mean score of the overall scale was 3.237 ($SD = 1.204$), slightly above 3, indicating that EFL learners' general acceptance of AI feedback for English writing was at a moderate level. Furthermore, the means of perceived usefulness (Mean = 3.157, $SD = 1.167$), perceived ease of use (Mean = 3.431, $SD = 0.853$), attitude toward use (Mean = 3.224, $SD = 1.212$), and behavioral intention (Mean = 3.236, $SD = 1.131$) all exceeded 3. However, with standard deviations surpassing 1 for all variables except perceived ease of use, the data indicate significant variability in learners' perceptions. This divergence implies that while some learners recognized the value of AI feedback for enhancing writing skills, others remained skeptical or unconvinced.

Table 5. Descriptive Statistics of Learners' Acceptance of AI Feedback

Dimensions	Mean	SD	Skewness ($SE = .155$)	Kurtosis ($SE = .309$)
Perceived Usefulness	3.157	1.167	-0.527	-0.377
Perceived Ease of Use	3.431	0.853	-0.356	0.265
Attitude Towards Use	3.224	1.212	0.667	0.372
Behavioral Intention	3.236	1.131	0.109	-0.103
Overall Scale	3.237	1.204	-0.241	-0.508

The interview results aligned with and further substantiated the questionnaire survey results. A majority of students acknowledged the potential utility of AI feedback in supporting the English writing process. However, mirroring the survey data, the interviews revealed significant disparities in participants' attitudes toward AI feedback, particularly across different proficiency levels. Notably, students with higher and intermediate writing proficiency expressed cautious concerns regarding Kimi's feedback. While they recognized its advantages—such as timely error detection and broad linguistic coverage—they critiqued its inability to interpret contextual nuances or the writer's underlying intent. These participants characterized AI feedback as overly generic and formulaic, contrasting it with the nuanced and personalized guidance teachers can offer. The following interview excerpts illustrate these critiques:

Excerpt 1: I have to admit that Kimi's feedback is useful for writing improvement, but it feels too robotic, as if it's churned out from templates. It lacks personal touch and detail, not as flexible or thorough as feedback from my English teacher (Student 1).

Excerpt 2: Although Kimi's feedback has some value, I prefer teacher feedback. My English teacher not only gives clear directions for improving my essay but also accurately identifies the specific issues (Student 6).

Excerpt 3: Kimi's feedback feels impersonal, unlike teacher feedback which feels warm. My English teacher often adds encouraging notes, which motivates me to write more (Student 11).

Conversely, students with lower writing skills depend more on AI feedback, appreciating its immediate guidance for quick learning and improvement. As evidenced by the excerpts below:

Excerpt 4: I have used other AI tools for Chinese writing. They can spot mistakes immediately and offer correct examples... In a word, I think their feedback is faster and more effective than teacher feedback (Student 5).

Excerpt 5: I find AI feedback very practical; it provides authentic vocabulary and syntax and assesses style appropriateness (Student 7).

Excerpt 6: Kimi's feedback is more comprehensive than teacher feedback. For me, the polished essay Kimi offered is the most valuable, since I can be exposed to authentic English expressions (Student 12).

Most students in the interviews advocate for a synergistic infusion of AI feedback and teacher feedback in writing instruction. This hybrid approach, as illustrated in the following excerpts, was perceived to mitigate the limitations of each feedback while capitalizing on their respective advantages:

Excerpt 7: I think AI tools can be used to provide feedback on the surface-level errors in the essays such as issues in lexical choice and grammatical accuracy, and then teachers are supposed to offer feedback on deeper-level guidance (Student 3).

Excerpt 8: I think English teachers can integrate AI feedback into their instruction to provide more targeted, data-driven, and comprehensive guidance for writing improvement (Student 4).

In conclusion, students with varying writing abilities have distinct perspectives and needs regarding AI feedback and teacher feedback. Those with higher and average abilities value personalized teacher feedback more highly, while those with lower abilities find immediate AI feedback more beneficial. The majority believe a combination of both can significantly enhance writing skills. These insights suggest teachers should consider integrating these feedback types to cater to diverse student needs in future writing instruction.

Discussion

Through a quasi-experimental design, this study investigated the impact of teacher feedback and feedback from the GenAI tool Kimi on the writing proficiency of university-level EFL learners. It also delved into students' acceptance and perceptions of Kimi's feedback through a questionnaire survey and semi-structured interviews.

The findings from the paired samples t-test reveal substantial enhancements in writing proficiency for both the experimental and control groups, suggesting that both types of feedback exerted a positive impact on writing skills. Notably, the post-test mean difference between the two classes was more pronounced than that observed in the pretest. However, the independent samples t-test and the repeated measures ANOVA indicated that the group variable did not significantly affect writing scores, which implies that there was no evident advantage of Kimi's feedback over teacher feedback in terms of elevating writing proficiency. This suggests that the mode of feedback may not be a pivotal determinant in improving students' writing competencies. It is important to note, however, that this finding may be influenced by several key factors. First, the relatively short intervention period of ten weeks may not have been sufficient to fully capture the nuanced differences in learning outcomes that could emerge over a longer trajectory, particularly with regard to the internalization and transfer of language skills. Second, the participants' intermediate level of English proficiency may have constrained their capacity to effectively process, apply, and internalize the feedback—regardless of its source (AI or human). This limitation could have, in turn, dampened the overall impact of the feedback on writing development. Third, the present study confined its writing tasks to TEM-4 argumentative essays. Different task types—such as expository, narrative, or academic writing—may elicit varied responses to feedback interventions. To build on these insights, future research should consider extending the intervention timeline, employing more refined measures of language competence, and comparing the effect of distinct task types on intervention. This would allow for a deeper understanding of how different feedback modalities influence various dimensions of writing proficiency, particularly as learners progress and develop greater linguistic awareness. The results align partially with the research conducted by (Wang & Zhang, 2025), who reported that participants experienced a significant increase in overall writing proficiency, as well as improvements in lexical and syntactic accuracy and complexity, following exposure to AI-provided feedback. It is important to highlight that the study by (Wang & Zhang, 2025) study lacked a control group, thus caution should be exercised when drawing comparisons with the present study. Although there were no significant differences between AI feedback and teacher feedback in shaping EFL learners' writing proficiency, AI feedback holds potential advantages in augmenting the efficiency of teachers' essay grading (Deng & Lin, 2022). Utilizing GenAI tools can substantially decrease the time teachers spend on reviewing and responding to student assignments, freeing up more time for other pedagogical activities (Elsaiary, 2025). This is particularly beneficial in large classes where providing personalized feedback can be challenging and taxing (Bai & Hu, 2017).

The results of the questionnaire survey indicate that learners generally accepted AI feedback, albeit with significant variation among individuals ($SD > 1$ in most dimensions). Such variability may be attributed to a range of individual difference factors, including learners' prior experiences with AI interactions, their writing proficiency, attitudes toward technology, and demographic characteristics such as gender, age, and educational background. While statistical analyses confirm the existence of variations among learners, our study did not explore the potential underlying causes, thereby overlooking an important opportunity to understand the nuances in how learners accept AI-generated feedback. Future research should consider integrating these potential moderating variables into the analytical framework to identify the key

determinants that shape learners' receptiveness to AI. The semi-structured interviews corroborated the survey findings, showing that most respondents, especially those with higher and average English writing skills, preferred feedback from teachers. In contrast, students with lower proficiency were more inclined towards AI feedback. Those who favored teacher feedback contended that it is more thorough, offers clearer guidance, and provides emotional support, which significantly enhances their motivation to write. On the other hand, those who preferred Kimi feedback believe it is more beneficial for writing improvement due to its clarity and specificity. This aligns with the finding of (Dai et al., 2023), who found that AI feedback is more detailed and readable than teacher feedback. It is important to note that students in this study received Kimi feedback via email and could not ask follow-up questions, a measure taken to prevent misuse of GenAI for writing assignments. Admittedly, allowing students to ask follow-up questions to Kimi might increase their preference for GenAI feedback. Furthermore, while AI writing feedback is recognized as a supplement to, rather than a substitute for, teacher feedback (Godwin-Jones, 2022; Wang et al., 2025), the majority of students are convinced that an integrated approach, combining teacher feedback with AI suggestions, is optimal for enhancing their writing abilities. This not only helps learners in quickly identifying and correcting errors but also greatly assists teachers in writing instruction. Enhancing students' capacities to generate, analyze, and apply feedback from GenAI addresses the inherent delays in traditional writing instruction associated with teacher feedback. Furthermore, it fosters the "human-AI collaboration" learning paradigm that underscores the synergy between human and AI technology (Pan & Niu, 2025).

Given predictions by AI industry leaders that everyday-capable GenAI technology is on the horizon and super artificial intelligence may emerge in the years ahead (Leonid et al., 2024), it is crucial for language teachers to gain a deep understanding of this rapidly evolving technology, its potential applications, and its profound implications for future language education. Currently, GenAI tools are not optimized for automated writing evaluation services, making it difficult to implement common text annotation features found in existing AWE programs within these tools. However, some mature AWE programs, such as GrammarlyGO, have begun integrating generative pre-trained transformer model technology. It is anticipated that in the near future, finely tuned and optimized LLMs will be specifically applied to language learning and teaching, including writing evaluation. As public access to these models increases, GenAI is likely to be integrated into widely used learning management systems within educational institutions, becoming an even more critical component of instructors' and students' practical activities.

4. CONCLUSION

This research compared the efficacy of writing feedback offered by the GenAI tool Kimi versus teacher feedback in enhancing the writing proficiency of EFL learners. It also sought to gather students' perspectives on receiving GenAI feedback. The findings revealed no significant disparity in the impact of these two feedback modalities on writing proficiency enhancement. Moreover, students were of the opinion that an amalgamation of both feedback types was optimal for bolstering their English writing abilities. The findings underscore the pedagogical value of strategically integrating both AI-generated and teacher-provided feedback within contemporary EFL writing instruction.

This study reveals key pedagogical insights: AI feedback and teacher feedback exhibit complementary strengths, necessitating seamless integration for EFL writing instruction. To this end, several strategies can be implemented: Firstly, teachers may employ AI tools as a preliminary assessment and feedback mechanism. By leveraging AI's rapid analysis capabilities, students' compositions can be promptly evaluated for grammar, spelling, and stylistic issues. This not only increases the efficiency of formative assessment but also provides learners with immediate, data-driven insights into their writing performance. However, to ensure this integration is meaningful and sustainable, institutional frameworks must be developed that support the responsible use of AI in writing instruction. Secondly, curriculum design should explicitly incorporate AI literacy and writing technologies as core components. Writing courses can include modules on interpreting AI feedback, understanding its limitations, and integrating it with human judgment. For example, students can be taught how to compare AI-generated suggestions with peer or teacher feedback, fostering critical engagement with technology. Additionally, interdisciplinary collaboration between language teachers and AI developers can help tailor tools to better align with pedagogical goals and linguistic diversity. Thirdly, training programs must evolve to prepare teachers for the AI-enhanced classroom. Professional development should include workshops on AI-assisted writing instruction, ethical considerations in AI use, and strategies for blending automated and human feedback. Teachers should be equipped not only with technical skills but also with the pedagogical knowledge to guide students in using AI as a supportive, rather than substitutive, tool. Institutions can also establish mentorship programs or learning communities where teachers share best practices and reflect on AI's impact on writing instruction. Fourthly, assessment policies need to be re-evaluated to accommodate AI's

role in writing instruction. While AI can streamline the evaluation of technical aspects, such as grammar and coherence, human assessment remains essential for evaluating creativity, argumentation, and critical thinking. A hybrid assessment model—combining AI-based formative feedback with teacher-led summative evaluation—can provide a balanced approach. Furthermore, grading rubrics should be updated to reflect new writing practices shaped by technology, ensuring that students are assessed fairly and holistically. Lastly, while AI offers transformative potential, ethical concerns must be addressed proactively. Over-reliance on AI may diminish students' ability to think independently and self-edit effectively. There is also a risk of plagiarism when students use AI to generate content without proper attribution. To mitigate these issues, institutions should incorporate lessons on academic integrity and responsible AI use into the curriculum. Additionally, efforts must be made to preserve the authenticity of student writing—AI should assist rather than replace the writing process. Technologies that track the evolution of student drafts and highlight AI contributions can help maintain transparency and accountability.

Despite these implications, this study is not without limitations. First, the relatively small sample size may restrict the generalizability of the findings. Second, the experimental design did not include a comparative group receiving combined feedback from both an AI tool and a writing instructor, which could have provided further insights into the efficacy of different feedback modalities. Third, the duration of the teaching intervention was limited to 10 weeks, leaving open the question of whether extended exposure would yield divergent outcomes. In light of these methodological constraints, future research should consider expanding the sample size, incorporating additional experimental conditions (e.g., no-feedback control groups and hybrid AI-instructor feedback groups), and extending the duration of the pedagogical intervention.

5. ACKNOWLEDGE

This article represents the interim research outcomes of the 2024 research project of the Sichuan Provincial Private Education Association titled “The Impact of Artificial Intelligence Technology on the Developmental Trajectories of College Students' English Writing Proficiency from the Perspective of the Complex Dynamic Systems Theory” (Grant No.: MBXH24YB191) and the 2025 “Three-Campus Initiative” teaching reform program of Geely University of China titled “Study on Human-AI Collaborative English Writing Assessment Model Driven by Artificial Intelligence” (Grant No.: 2025SGXYJG133).

6. REFERENCES

- Amani, N., & Bisriyah, M. (2025). University Students' Perceptions of AI-Assisted Writing Tools in Supporting Self-Regulated Writing Practices. *IJELTAL: Indonesian Journal of English Language Teaching and Applied Linguistics*, 10(1), 91–107. <https://ijeltal.org/index.php/ijeltal/article/view/1942>
- Bai, L., & Hu, G. (2017). In the Face of Fallible AWE Feedback: How do Students Respond?. *Educational Psychology*, 37(1), 67–81. <https://doi.org/10.1080/01443410.2016.1223275>
- Bai, L., & Wei, Y. (2024). Exploring EFL Learners' Integration and Perceptions of ChatGPT's Text Revisions: A Three-Stage Writing Task Study. *IEEE Transactions on Learning Technologies*, 17, 2161–2172. <https://doi.org/10.1109/TLT.2024.3491864>
- Chen, X. R. (2024). ChatGPT and Second Language Pragmatic Teaching. *Foreign Language Learning Theory and Practice*, 4, 16–29. <https://doi.org/10.13564/j.cnki.issn.1672-9382.2023.03.007>
- Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic Literature Review on Opportunities, Challenges, and Future Research Recommendations of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, 4, 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gašević, D., & Chen, G. (2023, July). Can Large Language Models Provide Feedback to Students? A Case Study on ChatGPT. In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)* (pp. 323–325). IEEE. <https://doi.org/10.1109/ICALT58122.2023.00100>
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319–340. <https://doi.org/10.2307/249008>
- Deng, J., & Lin, Y. (2022). The Benefits and Challenges of ChatGPT: An Overview. *Frontiers in Computing and Intelligent Systems*, 2(2), 81–83. <https://pdfs.semanticscholar.org/2bf0/2911cbc3bc24e43a910477a90363225bed95.pdf>
- 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>
- Dong, L. (2024). “Brave New World” or Not?: A Mixed-Methods Study of the Relationship Between Second Language Writing Learners’ Perceptions of ChatGPT, Behaviors of Using ChatGPT, and Writing Proficiency. *Current Psychology*, 43(21), 19481–19495. <http://doi.org/10.1007/s12144-024-05728-9>
- Elsaiary, R. (2025). Transforming Best Practices and Empowering Educators. In A. ElSayary (Ed.), *Prompt Engineering and Generative AI Applications for Teaching and Learning* (pp. 161–172). IGI Global Scientific Publishing. <http://doi.org/10.4018/979-8-3693-7332-3.ch010>
- Erisyerico, M. L., & Fauzan, A. (2024). Comparing Students’ Perceptions of AI and Human Feedback in Improving Writing Skills. *Scientia*, 3(2), 381–387. <http://doi.org/10.51773/sssh.v3i2.331>
- Feng, L. (2025). Investigating the Effects of Artificial Intelligence-Assisted Language Learning Strategies on Cognitive Load and Learning Outcomes: A Comparative Study. *Journal of Educational Computing Research*, 62(8), 1961–1994. <http://doi.org/10.1177/07356331241268349>
- Godwin-Jones, R. (2022). Partnering with AI: Intelligent Writing Assistance and Instructed Language Learning. *Language Learning & Technology*, 26(2), 5–24. <http://doi.org/10.125/73474>
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical Versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment*, 12(1), 19–30. <https://doi.org/10.1037/1040-3590.12.1.19>
- Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A Large-Scale Comparison of Human-Written Versus ChatGPT-Generated Essays. *Scientific Reports*, 13(1), 18617. <https://doi.org/10.1038/s41598-023-45644-9>
- Huawei, S., & Aryadoust, V. (2023). A Systematic Review of Automated Writing Evaluation Systems. *Education and Information Technologies*, 28(1), 771–795. <https://doi.org/10.1007/s10639-022-11200-7>
- Hyland, K. (2013). Student Perceptions of Hidden Messages in Teacher Written Feedback. *Studies in Educational Evaluation*, 39(3), 180–187. <https://doi.org/10.1016/j.stueduc.2013.06.003>
- Ingle, S. J., & Pack, A. (2023). Leveraging AI Tools to Develop the Writer Rather than the Writing. *Trends in Ecology & Evolution*, 38(9), 785–787. <https://doi.org/10.1016/j.tree.2023.05.007>
- Kline, R. B. (2023). *Principles and Practice of Structural Equation Modeling*. Guilford Publications.
- Kumar, R. (2023). Faculty Members’ Use of Artificial Intelligence to Grade Student Papers: A Case of Implications. *International Journal for Educational Integrity*, 19(1), 9. <https://doi.org/10.1007/s40979-023-00130-7>
- Leonid, G., Anton, G., & Igor, G. (2024). The Evolution of Artificial Intelligence: From Assistance to Super Mind of Artificial General Intelligence? Article 1. Information Technology and Artificial Intelligence: The Past, Present and Some Forecasts. *Social Evolution & History*, 23 (1), 156–183. <https://doi.org/10.30884/seh/2024.01.07>
- Li, K. (2023). Determinants of college students’ actual use of AI-based systems: An extension of the technology acceptance model. *Sustainability*, 15(6), 5221. <https://doi.org/10.3390/su15065221>
- Liu, G., & Ma, C. (2024). Measuring EFL Learners’ Use of ChatGPT in Informal Digital Learning of English Based on the Technology Acceptance Model. *Innovation in Language Learning and Teaching*, 18(2), 125–138. <https://doi.org/10.1080/17501229.2023.2240316>
- Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners’ critical thinking in EFL education? A systematic review. *Smart Learning Environments*, 12(1), 48. <https://doi.org/10.1186/s40561-025-00406-0>
- Mizumoto, A., & Eguchi, M. (2023). Exploring the Potential of Using an AI Language Model for Automated Essay Scoring. *Research Methods in Applied Linguistics*, 2(2), 100050. <https://doi.org/10.1016/j.rmal.2023.100050>
- Pan, H. Y., & Niu, B. G. (2025). Surveying GenAI-Enhanced Foreign Language Learners’ Writing Feedback Literacy. *Technology Enhanced Foreign Language Education*, 1, 94–101, 125. <http://doi.org/10.20139/j.issn.1001-5795.20250113>
- Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, Expectations, and Experience: Sources of Student Perceptions of Automated Writing Evaluation. *Computers in Human Behavior*, 70, 207–221. <http://doi.org/10.1016/j.chb.2016.12.076>
- Shermis, M. D., Burstein, J., & Bursky, S. A. (2013). Introduction to Automated Essay Evaluation In M. D. Shermis & J. Burstein (Eds.), *The Handbook of Automated Essay Evaluation: Current Applications and New Directions* (pp. 1–15). Routledge.

- Tate, T., Doroudi, S., Ritchie, D., Xu, Y., & Warschauer, M. (2023). Educational Research and AI-Generated Writing: Confronting the Coming Tsunami. *EdArXiv Preprints*. <https://doi.org/10.35542/osf.io/4mec3>
- Tseng, W., & Warschauer, M. (2023). AI-Writing Tools in Education: If You Can't Beat Them, Join Them. *Journal of China Computer-Assisted Language Learning*, 3(2), 258–262. <https://doi.org/10.1515/jccall-2023-0008>
- Wang, Y. B., Xuan, Q. Y., & Zhang, J. (2025). The Impact of Generative AI Intervention on Chinese EFL Students' Engagement with Writing Feedback. *Foreign Languages in China*, 22(2), 71–78. <http://doi.org/10.13564/j.cnki.issn.1672-9382.2025.02.009>
- Wang, J., & Zhang, X. (2025). Exploring the Role of AI Technology in Shaping College Students' English Writing Development: Insights from the Complex Dynamic Systems Theory. *Journal of Azerbaijan Language and Education Studies*, 2(3), 31-46. <https://doi.org/10.69760/jales.2025002012>
- Wei, S., & Li, L. Y. (2023). Artificial Intelligence-Assisted Second Language Writing Feedback: A Case Study of ChatGPT. *Foreign Languages in China*, 20(3): 33–40. <https://doi.org/10.13564/j.cnki.issn.1672-9382.2023.03.007>
- Weigle, S. C. (2013). English as a second language writing and automated essay evaluation. In M. D. Shermis & J. Burstein (Eds.), *The Handbook of Automated Essay Evaluation: Current Applications and New Directions* (pp. 36–54). Routledge.
- Zhao, Y. (2015). *Second Language Creative Writers: Identities and Writing Processes*. Multilingual Matters.
- Zheng, R. (2025). AWE and LLMs in L2 writing feedback: An exploration of EFL learners' experiences. *Academic Journal of Humanities & Social Sciences*, 8(7), 60-69. <https://doi.org/10.25236/AJHSS.2025.080708>