



Integrating teacher and AI-generated feedback to enhance EFL academic writing: A mixed-methods study of performance and perceptions

ABSTRACT - Academic writing presents considerable challenges for English as a Foreign Language (EFL) learners, particularly in mastering higher-order skills such as idea development and organization, as well as lower-order skills including vocabulary, grammar, and writing mechanics. Although instructor feedback provides valuable contextual and rhetorical support, it is frequently constrained by limited time and heavy workloads. In contrast, AI-based feedback offers prompt and detailed language corrections but lacks the capacity for in-depth conceptual and rhetorical evaluation. This study investigates the impact of combining lecturer and AI-generated feedback on EFL students' academic writing achievement and examines their perceptions of this hybrid approach. An explanatory sequential mixed-methods design was implemented with 25 fifth-semester EFL students. Participants completed a pre-test, participated in an integrated feedback intervention, took a post-test, and subsequently responded to a questionnaire and semi-structured interviews. Quantitative findings revealed a statistically significant increase in overall writing scores from 67.56 to 79.48 ($p < 0.001$), with a very large effect size (Cohen's $d = 1.78$). Improvements were evident across all analytic writing components. Qualitative results indicated positive student perceptions, with lecturer feedback identified as conceptually clarifying and AI feedback as technically beneficial. Overall, the integrated feedback model effectively strengthened both linguistic accuracy and rhetorical development in EFL academic writing.

Putri Hindun Anifayakun Shahab^{*1}

Irwandi²

Ilham³

^{1,2,3}University of Muhammadiyah
Mataram, Indonesia

^{*}Corresponding email;

putrishahab97@gmail.com

Article History

Submitted: 4 March 2026

Accepted: 14 May 2026

Published: 31 May 2026

Keywords

Teacher feedback; AI-generated feedback; Academic writing; Human-AI collaboration; EFL learners

Recommended APA Citation

Shahab, P. H. A., Irwandi., & Ilham. (2026). Integrating teacher and AI-generated feedback to enhance EFL academic writing: A mixed-methods study of performance and perceptions. *Englisia: Journal of Language, Education, and Humanities*, 13(2), 511-531.

<https://doi.org/10.22373/englisia.72>

1. Introduction

The rapid proliferation of generative AI in education has fundamentally altered how students receive writing feedback, yet little is known about how to optimally integrate it with teacher input. In the context of higher education in various countries, writing competence is recognized as an important prerequisite for optimal academic participation as well as the development of knowledge in a specific field of study (Cheng & Zhang, 2021; Steiss et al., 2024; Yang et al., 2023). Although its urgency has been widely emphasized, the reality is that many students continue to experience difficulties in composing texts that are systematically organized and meet academic conventions. This condition indicates the need for learning strategies that can provide support consistently, purposefully, and sustainably to improve the quality of students' academic writing. Teacher feedback serves as a pivotal mechanism in guiding students toward enhanced writing outcomes, particularly by facilitating the refinement of content development, the strengthening of rhetorical organization, and the bolstering of argumentation skills. Empirical studies consistently demonstrate that teacher feedback enables learners to comprehend expectations and implement more strategic revisions to their drafts (Cheng & Zhang, 2021; Do & Phan, n.d.; Lo et al., 2025; Yang et al., 2023). However, teacher feedback is frequently constrained by temporal limitations, workload demands, and the necessity to deliver individualized instruction to sizable student populations. These constraints diminish both the frequency and depth of feedback provided, thereby highlighting the imperative for supplementary support mechanisms that can operate concurrently with teacher guidance to address persistent challenges in student writing.

Striving to develop higher-order academic writing competencies. Numerous EFL learners encounter substantial difficulties in constructing coherent arguments, selecting appropriate academic vocabulary, maintaining syntactic accuracy, and ensuring logical progression across paragraphs (Baroudi et al., 2025; Song & Song, 2023; Weber et al., 2025; Wilson et al., 2024). These multidimensional challenges frequently persist despite formal instruction, suggesting that learners require feedback that is more continuous, scaffolded, and individualized than what educators can feasibly provide independently. As these challenges continue to intensify, the necessity for supplementary feedback systems becomes increasingly evident, thereby directing attention toward technology-based writing support.

Revision processes. Such technological instruments can assist learners in enhancing grammatical precision, broadening lexical repertoire, and improving clarity at the sentence level (Dizon & Gold, 2023; Escalante et al., 2023; Fleckenstein et al., 2023; Guan et al., 2024; Huang et al., 2024). Despite these considerable advantages, AI systems remain constrained in their capacity to evaluate contextual meaning, the cogency of arguments, or dimensions of rhetorical appropriateness, whereby human educators continue to fulfil an essential interpretive function (Steiss et al., 2024). These complementary strengths and limitations demonstrate that feedback derived exclusively from either teachers or AI systems cannot adequately address the multifaceted requirements of academic writing. Consequently, the integration of both feedback modalities has emerged as a promising pedagogical approach in writing instruction.

Prior studies have explored teacher feedback, automated writing evaluation, and technology-enhanced writing approaches, reporting positive effects on linguistic accuracy, learner motivation, engagement, and drafting practices (Baroudi et al., 2025; Guan et al., 2024; Lo et al., 2025; Song & Song, 2023; Weber et al., 2025; Wilson et al., 2024). Comparative investigations of human and AI-generated feedback further indicate that each source provides distinct advantages, depending on the specific writing dimension under assessment (Dizon & Gold, 2023; Escalante et al., 2023; Fleckenstein et al., 2023; Huang et al., 2024; Lo et al., 2025; Steiss et al., 2024). Nevertheless, empirical research focusing on the integration of teacher and AI feedback remains limited. Existing studies tend to examine either learners' perceptions or discrete linguistic outcomes in isolation. Few investigations have addressed multidimensional improvements based on analytic rubrics or triangulated performance data with students' perceptions. This represents a significant gap in the literature, considering that learner engagement and interpretation substantially influence the effectiveness of feedback practices.

This gap emphasises the importance and originality of this study. First, there is still little empirical evidence on how the simultaneous integration of teacher feedback and AI-generated feedback affects various analytical dimensions of writing, including content, organization, vocabulary, language use, and mechanics. Second, attention to students' perceptions of hybrid feedback practices remains limited, even though previous studies have shown that learner engagement plays a crucial mediating role in determining learning outcomes (Baroudi et al. 2025; Ranalli 2021; Song & Song 2023). Third, there are still relatively few studies that systematically combine quantitative measures of writing improvement with qualitative insights into student experiences, particularly in the field of academic writing in English as a foreign language (EFL) and especially in the context of Indonesian higher education.

Therefore, the aim of this study is to investigate the impact of integrating teacher and AI-generated feedback on EFL students' academic writing performance. Specifically, the study addresses three research questions:

1. To what extent does the integration of teacher and AI-generated feedback improve students' overall academic writing performance?
2. How does integrated feedback influence each analytic component of the writing rubric?
3. What are students' perceptions of the integrated feedback process?

2. Literature review

Literature review is a comprehensive investigation of the available theoretical background including from books and scholarly articles related to your research areas and theories. In this section, you should provide a description, summary, and critical evaluation of your works concerning the research problem being investigated. Literature reviews are aimed at providing an overview of sources you have explored while researching a particular topic to notify your readers how your research fits within a larger field of study.

The landscape of writing instruction has evolved further through the integration of generative AI tools. Barrot, (2023) emphasizes that the use of ChatGPT in second language (L2) writing instruction offers substantial potential for direct, targeted support that addresses learners'

linguistic challenges. However, the effectiveness of this technology depends on the dynamics of learner interaction. Koltovskaia, (2023) indicates that student engagement with automated writing evaluation is a determining factor in the tangible improvement resulting from such feedback. Furthermore, regarding the dichotomy between automated and manual interventions, Mizumoto & Eguchi, (2023) conclude that while AI feedback is efficient, its nature differs from human feedback, making a hybrid approach the most optimal for language development.

2.1. Writing

Academic writing in higher education constitutes a complex and indispensable competency that functions as a form of social communication, enabling writers to participate substantively in their academic and professional communities (Hyland, 2019). This competency encompasses mastery of discipline-specific conventions as well as rhetorical structures that are crucial for understanding scholarly discourse (Swales, 2012).

Proficiency in academic writing is especially crucial for early-career researchers and college students. However, many students face significant challenges in producing coherent, logically structured, and rigorously analytical texts, indicating a persistent gap between institutional expectations and the reality of students' writing abilities. (Mendoza et al., 2025; Mishra, 2023).

Moreover, embedding academic writing instruction across disciplines and adopting innovative pedagogical approaches are essential for cultivating critical thinking and effective communication skills (Mendoza et al., 2025; Mishra, 2023). Consequently, systematic and interdisciplinary frameworks for teaching academic writing are imperative to develop competent writers who can successfully meet the intellectual and communicative demands of higher education (Coffin et al., 2005).

2.1.1. Challenges in EFL academic writing

The difficulties inherent in academic writing are particularly pronounced in English as a Foreign Language (EFL) context, where learners are required to navigate linguistic constraints alongside complex rhetorical expectations. Empirical studies indicate that EFL students often struggle to construct coherent arguments, sustain logical organization, and employ discipline-appropriate academic vocabulary (Cheng & Zhang 2021; Hyland 2020). These challenges are frequently linked to limited exposure to academic discourse conventions and insufficient opportunities for structured revision and feedback.

Moreover, persistent language-related issues including grammatical errors, restricted lexical range, and inadequate textual cohesion tend to diminish the overall quality and clarity of students' written work (Ramzan et al., 2023). Collectively, these multifaceted constraints highlight the necessity of comprehensive and systematic feedback practices that address both global concerns, such as argumentation and organization, and local concerns, such as grammar and vocabulary.

To design such feedback practices effectively, it is essential to first understand the specific dimensions of academic writing that determine overall text quality. Performance in academic writing is generally assessed using analytical scoring rubrics that evaluate various components. Content refers to the depth, relevance, and clarity of the ideas presented. Organization concerns the logical sequence and coherence between paragraphs. Vocabulary encompasses lexical appropriateness and variety, while language use addresses the accuracy of grammar and the complexity of syntax. Mechanics include punctuation, spelling, and formatting conventions (Jacobs, 1981; Weigle, 2002). These components are interdependent and together determine the overall quality of a written text. Effective written instructions therefore require attention both at the global level (content and organization) and at the local level (grammar and mechanics), as weaknesses in any dimension can reduce the effectiveness of communication (Ferris, 2011).

2.2. Teacher feedback in writing instruction

Teacher feedback has consistently been regarded as a fundamental component of writing pedagogy. It offers individualized support to students in developing ideas, organizing rhetorical structures, and refining arguments, thereby facilitating more meaningful and sustained revision processes (Ferris 2014; Hyland & Hyland 2006). Empirical evidence suggests that systematically delivered feedback enhances students' understanding of academic standards and fosters more purposeful and strategic revision practices.

Nevertheless, the impact of teacher feedback may be constrained by practical factors, including large class enrolments and substantial instructional workloads. Such limitations can diminish both the quantity and quality of feedback provided, while also restricting opportunities for continuous interaction and personalized guidance (Maier & Klotz, 2022). Consequently, supplementary feedback approaches have increasingly been investigated as potential means of strengthening students' writing development.

2.3. AI-generated feedback in writing

Recent developments in artificial intelligence have led to the emergence of automated writing evaluation systems capable of delivering immediate and detailed feedback. These technologies assist learners in identifying grammatical errors, refining lexical choices, and addressing sentence-level concerns (Ranalli 2021; Stevenson & Phakiti, 2014). The provision of timely AI-generated feedback facilitates iterative drafting and fosters greater learner autonomy in the revision process.

However, AI-based feedback remains limited in its capacity to evaluate contextual meaning, the coherence and persuasiveness of arguments, and rhetorical appropriateness. In contrast, human instructors retain a critical interpretive function in addressing complex writing issues that demand disciplinary expertise and contextual awareness (Wilson & Roscoe, 2020). Accordingly, AI-generated feedback is generally conceptualized as a complementary resource rather than a substitute for teacher guidance.

2.4. *Integrated teacher–AI feedback and research gaps*

Considering the respective strengths and limitations of teacher feedback and AI-generated feedback, their integration represents a potentially effective approach in writing instruction. A growing body of research on teacher feedback, automated writing evaluation, and technology-enhanced writing instruction has documented positive effects on linguistic accuracy, learner motivation, engagement, and drafting practices (Baroudi et al., 2025; Guan et al., 2024; Lo et al., 2025; Song & Song 2023; Weber et al., 2025; Wilson et al., 2024). Comparative studies examining human and AI feedback further indicate that each source provides distinct benefits, depending on the specific dimension of writing being evaluated (Dizon & Gold 2023; Escalante et al., 2023; Fleckenstein et al., 2023; Huang et al., 2024; Lo et al., 2025; Steiss et al., 2024).

Nevertheless, empirical investigations into fully integrated teacher–AI feedback models remain limited. Existing research has tended to focus primarily on learners’ perceptions or on discrete linguistic outcomes, with relatively little attention to writing development assessed through multidimensional rubrics. Moreover, few studies have combined quantitative measures of performance improvement with qualitative analyses of learner experiences, particularly within EFL contexts and higher education settings in Indonesia. These gaps underscore the need for comprehensive empirical research that simultaneously examines measurable gains in writing proficiency and students’ perceptions of hybrid feedback practices.

3. Method

3.1. *Research design*

This study employed an explanatory sequential mixed-method design, consisting of a quantitative phase followed by a qualitative phase. The quantitative phase implemented a one-group pretest–post-test quasi-experimental design to examine the impact of integrating teacher and AI-generated feedback on students’ academic writing performance. The qualitative phase involved semi-structured interviews that were used to further explain and elaborate on the quantitative findings. The participants of the study were 25 fifth-semester EFL students at X University who were at the A2 (Elementary) proficiency level, and were enrolled in an Academic Writing course. Fifth-semester students were intentionally selected because they had previously completed essential prerequisite courses such as basic writing and grammar. Thus, they possessed adequate linguistic competence and writing experience to meaningfully engage with both teacher-provided and AI-generated feedback. All participants voluntarily completed every stage of the study including the pretest, intervention, post-test, questionnaire, and interviews. This design and participant selection align with the purpose of the study which aimed to evaluate the effectiveness of the integrated feedback model and understand students’ perceptions toward the feedback process. Although this study adopted a single-group design without a control group due to curricular constraints, the triangulation of qualitative data and the substantial effect size ($d = 1.78$) provide a strong methodological foundation for the conclusions drawn.

3.2. Instruments

This study employed three main instruments to collect data: a writing test (pretest and post-test), questionnaires, and interviews.

3.2.1. Writing test

The primary instrument for measuring students' writing performance was an academic writing test administered before and after the intervention. Both pretest and post-test required students to produce an academic paragraph or short essay, assessed using an analytic scoring rubric adapted from Jacobs et al. (1981). The rubric evaluated five components: content, organization, vocabulary, language use, and mechanics. These components allowed a detailed analysis of the specific areas improved through integrated teacher and AI-generated feedback.

3.2.2. Intervention

The intervention adopts a three-stage structured feedback model. First, students receive automated feedback via ChatGPT (v. 4) to identify grammatical and structural deficiencies. Second, students revise their drafts based on these recommendations. Third, instructors provide personalized feedback to address more complex rhetorical issues. Finally, students conduct a comparative evaluation and final alignment to refine their academic essays.

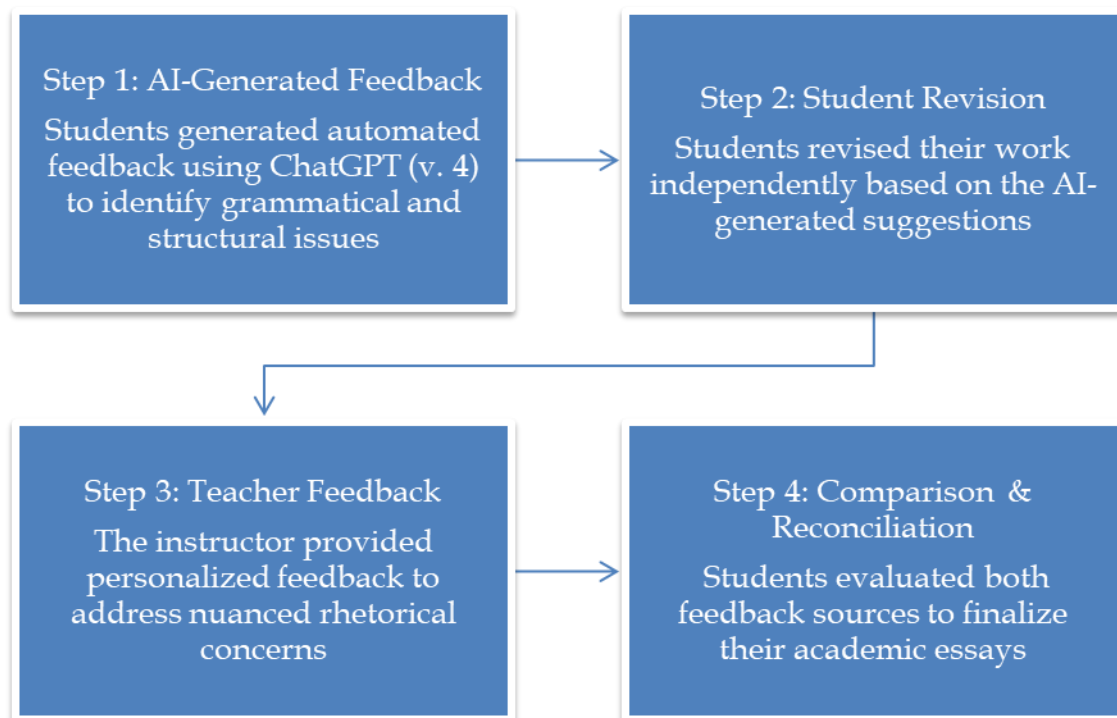


Figure 1. The integrated feedback procedure

3.3. Survey questionnaire

The perception questionnaire consisted of 10 Likert-scale items (1 = strongly disagree; 5 = strongly agree) designed to measure students' perceptions of the clarity, usefulness, accuracy, and effectiveness of the integrated feedback. Validity. Pearson item-total correlations showed

that all questionnaire items exceeded the minimum critical value ($r > .396$), indicating strong construct validity. Reliability. Cronbach's Alpha was calculated at .959, demonstrating excellent internal consistency and ensuring that the instrument reliably captured students' perceptions.

3.4. Interview

Semi-structured interviews were conducted with a purposively selected group of students to gather qualitative insights into their experiences with both teacher-provided and AI-generated feedback. The interview protocol explored participants' perceptions of writing improvement, the clarity and usefulness of the feedback received, and any challenges or limitations encountered during the feedback process. Each interview lasted approximately 10–15 minutes. These sessions were designed as brief check-ins to document spontaneous reflections, rather than in-depth phenomenological explorations, in order to minimize participant fatigue. Each session was audio-recorded with consent, and subsequently transcribed to facilitate systematic analysis.

3.4.1. Intervention

The study procedures comprised four sequential stages aligned with the explanatory sequential mixed-methods design. First, during the pretest administration stage, students completed a writing task that was scored using an analytic rubric to establish a baseline measure of their initial writing ability. Second, in the feedback integration stage, students received both teacher feedback—focusing on content and organization—and AI-generated feedback using ChatGPT-4, which emphasized vocabulary choice, grammatical accuracy, and mechanical conventions. Students subsequently revised their drafts by prioritizing instructors' input concerning content substance and presentation while leveraging AI to ensure linguistic precision. Third, a post-test administration was conducted using a parallel writing task, scored with the same analytic rubric to ensure consistency and to measure improvement following the intervention. Fourth, during the perception and reflection stage, students completed a survey questionnaire and participated in semi-structured interviews to reflect on their experiences with the integrated feedback model. Collectively, these four stages ensured that both quantitative and qualitative data were systematically collected, thereby facilitating a comprehensive evaluation of the hybrid feedback intervention's effectiveness on students' academic writing development.

3.5. Data analysis

The quantitative and qualitative data were analyzed sequentially following the explanatory mixed-method design. All quantitative analyses were conducted using SPSS. Before performing inferential tests, the Shapiro–Wilk test was used to examine the normality of pretest and post-test score distributions. To address RQ1, a paired sample t-test was employed to determine whether there was a significant difference between students' pretest and post-test overall writing scores. The magnitude of improvement was further examined using Cohen's d to determine the effect size of the intervention.

For RQ2, the five analytic writing components (content, organization, vocabulary, language use, and mechanics) were analyzed separately. Because the normality assumption was

not met for these component scores, the Wilcoxon Signed-Rank Test was used to evaluate whether significant improvements occurred across each writing aspect.

To answer RQ3, descriptive statistics including mean scores and standard deviations were calculated for each item of the student perception questionnaire. The descriptive results were used to determine the level of students' perceived usefulness and clarity of the integrated teacher–AI feedback. The validity of the questionnaire was confirmed through item–total correlations, and reliability was supported by a Cronbach's Alpha coefficient of .959, indicating excellent internal consistency.

The qualitative data from student interviews were analyzed using thematic analysis. Interview transcripts were coded to identify recurring patterns related to students' experiences with teacher and AI-generated feedback. These qualitative findings were integrated with the quantitative results to provide explanatory insights regarding the effectiveness of the intervention, consistent with the explanatory sequential mixed-method approach.

4. Findings

This section reports both the quantitative and qualitative findings in relation to the three research questions guiding the study. It presents the results of descriptive statistical analyses, normality testing, and inferential procedures, along with questionnaire data and complementary qualitative evidence to support the overall interpretation.

4.1. Improvement in students' writing performance

To provide a clear empirical basis for evaluating the effectiveness of the intervention, the analysis begins by examining the descriptive statistics of students' writing performance before and after the implementation of the hybrid feedback model.

Tabel 1

Descriptive statistics of pre-test and post-test score.

Variable	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Pre-test	25	36	89	67.6	15,171	230.2
Post-test	25	55	95	79.5	11.225	126.0

Table 1 summarizes the students' writing performance before and after the learning intervention, which combines feedback from teachers and feedback from artificial intelligence (AI). Before the intervention, the average score on the initial test was 67.6, with scores ranging from 36 to 89, indicating significant differences in the students' initial writing skills. The relatively large standard deviation (SD = 15.2) and variance (230.2) further indicate that the students' performance levels were widely dispersed.

After the intervention, the students' post-test results showed significant improvement, with the average score rising to 79.5 and the score range narrowing to 55 to 95. In addition, a lower standard deviation (SD = 11.2) and variance (126.0) indicated a more consistent level of

performance among students. These results suggest that the intervention not only improved the overall writing skills of participants, but also reduced the gap between low- and high-performing students.

Tabel 2

Test of normality (Shapiro-Wilk).

Variabel	N	Statistic	df	Sig.
Pretest	25	.955	25	.327
PostTest	25	.931	25	.090

Table 2 shows the results of the Shapiro-Wilk test for normal distribution, which was performed on the writing scores before and after the test. The Shapiro-Wilk method is generally recommended for small to medium sample sizes ($n < 50$) and is therefore suitable for evaluating the distribution of the 25 writing scores obtained in this study. The analysis shows that the pre-test results yielded a Shapiro-Wilk value of 0.955 with a significance of 0.327, while the post-test results yielded a value of 0.931 with a significance of 0.090. In both cases, the p-value is above the conventional significance threshold of 0.05, suggesting that both distributions do not deviate significantly from normality. These results show that the data adequately conform to a normal distribution and that the assumption of normality is satisfied. Ensuring normality is an essential prerequisite for the application of parametric statistical analyses. Since the pre- and post-test results satisfy this assumption, the use of a paired t-test is considered appropriate to test whether the integrated feedback intervention leads to a significant improvement in students' writing performance. If the normality assumption is not met, a nonparametric alternative such as the Wilcoxon rank sum test would be required. However, as shown in Table 2b, the results of the Shapiro-Wilk test for the writing component indicated that the assumption of normality was not met ($p < 0.05$). Therefore, a nonparametric analysis using the Wilcoxon Signed-Rank test was applied to evaluate student performance on each component.

Table 2b

Tests of Normality.

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre_Content	.146	25	.177	.914	25	.038
Post_Content	.144	25	.194	.908	25	.027
Pre_Org	.175	25	.048	.900	25	.019
Post_Org	.166	25	.074	.879	25	.007
Pre_Vocab	.225	25	.002	.911	25	.033
Post_Vocab	.193	25	.017	.844	25	.001
Pre_Lang	.127	25	.200*	.951	25	.263
Post_Lang	.204	25	.009	.915	25	.039
Pre_Mech	.210	25	.006	.925	25	.068
Post_Mech	.239	25	.001	.859	25	.003

Table 3

Paired samples t-test of pre-test and post-test scores.

Pair	Mean-Difference	Std. Deviation	Std. Error Mean	t	df	Sig. (2-Tailed)
Pretest-Posstest	-11.920	6.677	1.335	-8.927	24	< .001

Table 3 shows the results of the paired t-test used to assess significant differences in students' writing performance before and after the integrated feedback intervention. This analytical method was deemed appropriate due to the repeated measures design, with normality checked using the Shapiro-Wilk test. The analysis revealed an average difference of -11.920 between the pre- and post-test results, corresponding to an average improvement in post-test results of nearly 12 points; the negative sign corresponds to the calculation of pre-test results minus post-test results. Furthermore, the standard deviation of 6.677 indicates moderate variability in the degree of improvement among students, while the standard deviation of 1.335 represents a stable estimate for the average difference in relation to the sample size.

The inferential statistical analysis yielded a t-value of -8.927 with 24 degrees of freedom, resulting in a significance level of $p < 0.001$. This minimal p-value provides convincing statistical evidence that the observed improvement in writing skills is not a coincidence and confirms that the integrated feedback from teachers and AI has a significant positive impact on writing skills. Furthermore, the magnitude of the t-value underscores the consistency of this improvement across all participant groups. This striking t-value shows that the intervention enabled measurable progress for most students, suggesting broad effectiveness that benefits more than just a small portion of the class.

Table 4

Effect size (Cohen's d) for the difference between pre-test and post-test.

Statistic	Value	Interpretation
Mean Difference	11.920	Shows improvement between pre-test – post-test
Standard Deviation of Difference	6.677	variation in improvement
Cohen's d	1.78	Very large effect

Table 4 summarizes the analysis of effect size used to evaluate the extent of improvement associated with the integrated feedback intervention. An average difference of 11.920 was observed, corresponding to an average improvement of approximately twelve points from the pre-test to the post-test. This significant improvement demonstrates a substantial increase in writing skills achieved through the integration of feedback from teachers and feedback from artificial intelligence (AI).

The standard deviation of the point difference, which was reported as 6.677, describes the dispersion of individual performance improvement. Despite individual differences in progress, the consistency of positive results across the entire sample underscores the reliability of the

intervention. Cohen's d value is 1.78, which classifies the effect as very large. According to the conventional standard, where an effect size above 0.80 is considered large, this value indicates a very strong effect on writing performance. The results thus show that the observed improvements are not only statistically significant, as determined by the t -test, but also have practical significance in an educational context.

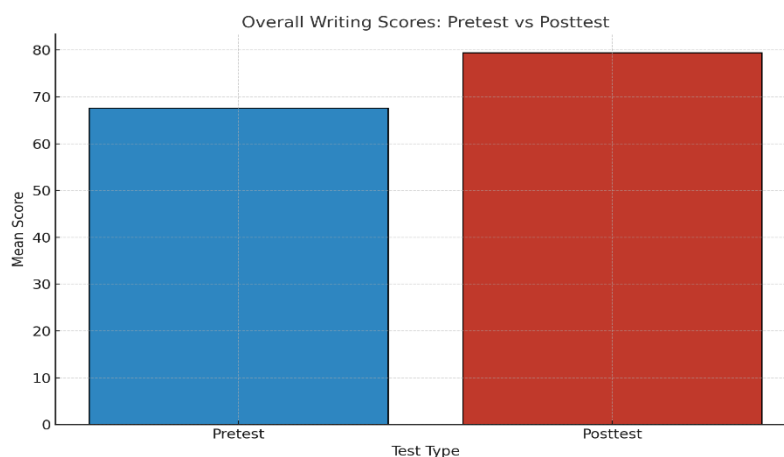


Figure 1. Comparison of overall pre-test and post-test writing scores

Figure 1 shows the average comparative values determined from the pre- and post-test assessments. Visual analysis of the data reveals a clear upward trend in writing skills following the implementation of the integrated feedback intervention. This graphical representation is consistent with the inferential statistics, which show that the observed improvement is statistically significant. Therefore, this visual evidence supports the quantitative results and reinforces the conclusion that the intervention led to a significant improvement in students' writing skills.

4.2. Component-level improvement (RQ2)

Table 5

Wilcoxon Signed-Rank test results for writing components.

Writing Component	Z Value	Sig. (2-tailed)	Interpretation
Content	-4.397	< .001	Significant Improvement
Organization	-4.286	< .001	Significant Improvement
Vocabulary	-4.311	< .001	Significant Improvement
Language Use	-4.172	< .001	Significant Improvement
Mechanics	-4.378	< .001	Significant Improvement

Since some writing component scores did not meet the assumption of normality, the Wilcoxon rank sum test was used to compare pre- and post-test scores for each writing component. The analysis revealed a consistent and statistically significant improvement in all five analytical components: content ($Z = -4.397$, $p < .001$),

organization ($Z = -4.286, p < .001$),
 vocabulary ($Z = -4.311, p < .001$),
 language use ($Z = -4.172, p < .001$), and
 mechanics ($Z = -4.378, p < .001$)

These results show a significant positive change in each component, suggesting that integrated feedback has a far-reaching impact on the development of students' writing. The improvements in content and organization show that students have improved the clarity of their ideas and the overall structure of their texts, while the improvements in vocabulary, language use, and mechanics reflect greater linguistic accuracy and control. Overall, these results show that the intervention contributes to the development of both higher-order and lower-order writing skills.

3D Chart of Pretest vs Posttest Across Writing Components (Soft Colors)

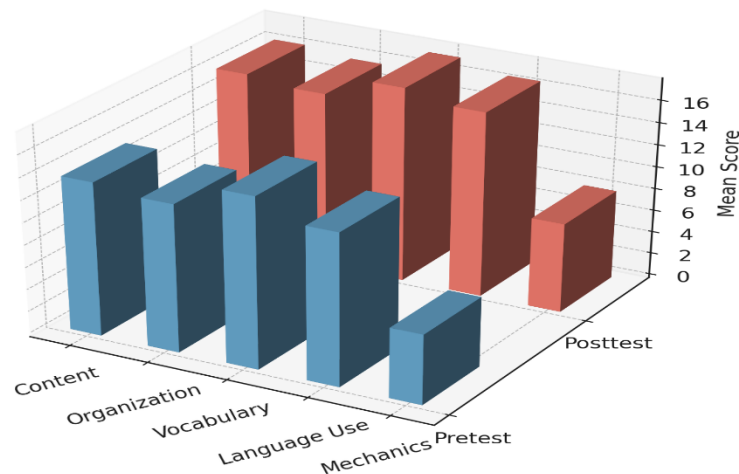


Figure 2. Comparison of pre-test and post-test scores across five writing components

Figure 2 shows the observed changes in the five components of writing: content, organization, vocabulary, language use, and mechanics. The figure shows consistent improvement in all components, with the improvement being most evident in language use and content. This visual trend supports the results of the Wilcoxon rank sum test, which shows a statistically significant improvement in each analytical category.

4.3. Students' perceptions of integrated teacher-AI feedback (RQ3)

Table 6 presents the descriptive statistics of students' perceptions toward the integrated use of teacher and AI-generated feedback in their academic writing. The mean scores for the ten questionnaire items ranged from 3.16 to 3.52, indicating generally moderate to high levels of positive perception.

Table 6

Descriptive statistics of questionnaire items.

Item	Mean	SD	Category
P01	3.44	1.121	High
P02	3.28	1.308	Moderate
P03	3.36	0.995	Moderate
P04	3.28	1.137	Moderate
P05	3.16	0.898	Moderate
P06	3.32	1.108	Moderate
P07	3.28	1.242	Moderate
P08	3.44	1.003	High
P09	3.52	1.159	High
P10	3.40	1.258	Moderate

Three specific points, identified by codes P01, P08, and P09, achieved average scores in the “high” category of 3.44, 3.44, and 3.52, respectively. These points relate to students' perceptions of the clarity of revisions (P01), improvement in the writing process (P08), and efficiency in identifying weaknesses in writing (P09). The high average scores show strong agreement among participants regarding the practical benefits of feedback from two complementary sources. The remaining survey items were classified as “moderate,” with average scores between 3.16 and 3.40.

These items assess dimensions such as confidence in the use of feedback, the perceived value of AI-generated suggestions, and support for content development and organization. Although these values are slightly lower than those in the high category, they still reflect an overall positive attitude. This shows that while students find this intervention useful, some areas may need further optimization, particularly with regard to the clarity of AI recommendations and students' ability to interpret automated feedback.

The standard deviation values range between 0.898 and 1.308, indicating moderate variability between responses. It is noteworthy that higher standard deviation values were observed for items P02 (SD = 1.308) and P07 (SD = 1.242), indicating greater diversity in perception in these specific areas. This dispersion shows that individual students have different comfort levels and different results when it comes to integrating digital and teacher-led feedback mechanisms.

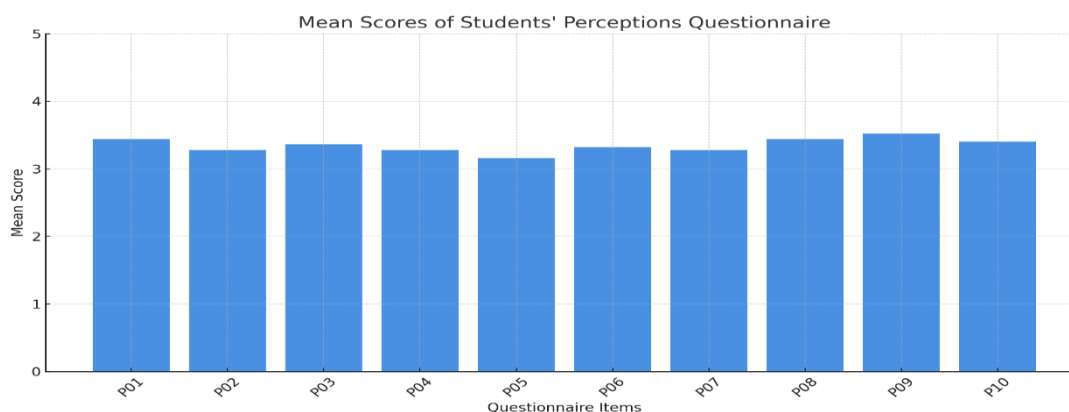


Figure 3. Mean scores of students' perceptions questionnaire

Figure 3 shows the average values determined from the ten questions in the questionnaire used to assess students' perceptions of the integrated feedback model. The distribution of values is generally in the medium to high range, indicating positive acceptance of the combined feedback approach between teachers and artificial intelligence (AI).

4.3. Interview results

Qualitative analysis of the interview transcripts revealed several recurring themes that elucidate student experiences with the integration of teacher and AI-generated feedback. These insights serve to contextualize the questionnaire findings, providing a deeper understanding of the factors contributing to the generally positive reception of the integrated feedback model.

4.3.1. Complementary strengths of teacher and AI feedback

The analysis of interview data shows that participants consistently view feedback from teachers and AI as having different levels of importance for the writing process. Feedback from teachers is considered important for explaining conceptual ideas, improving paragraph structure, and ensuring academic appropriateness. AI feedback, on the other hand, is valued for providing quick and detailed corrections related to grammar, vocabulary, and sentence clarity. Qualitative evidence supports this view, with one participant explaining that “The teacher helped explain the flow of ideas and identified areas requiring improvement, whereas the AI directly highlighted grammatical errors and unclear sentence constructions.” Another participant emphasized the importance of both sources, noting, “Relying on one source is not enough; the teacher focuses on content, while AI focuses on technical aspects.” These statements show that students view the integrated model as more comprehensive than relying on a single source of feedback.

4.3.2. Increased awareness of writing weaknesses

Many students report that hybrid feedback has heightened their awareness of their individual weaknesses in writing. The AI-generated feedback helps them identify recurring grammatical errors and repetitive or unclear sentence structures, while the teachers' feedback highlights more profound issues such as weak transitions and insufficiently developed arguments. One participant commented “The AI made me aware that I frequently used the same words repeatedly,” while another said, “The teacher typically pointed out when my paragraphs jumped between ideas or when my arguments lacked sufficient strength.” This increased awareness allows students to revise their writing more carefully and helps them understand the root causes of their writing difficulties.

4.3.3. Enhanced confidence and motivation

The students reported that they felt more confident and motivated during the revision process thanks to the support they received from two different perspectives. The feedback generated by artificial intelligence (AI) promoted a sense of autonomy and enabled students to revise their drafts independently before consulting with their teachers, while the feedback from teachers provided certainty and clarity in context. Qualitative evidence supports this view. One

of the participants explained: “Initially, I felt confused, but over time I became more confident because I knew what needed to be improved.” Another participant commented, “Receiving two types of feedback simultaneously was indeed substantial, but it ultimately made me more confident in my revisions.” This sense of guidance reduces anxiety and motivates students to become more involved in the revision process.

4.3.4. Cognitive overload and selective use of feedback

Although most feedback was positive, some students said they felt overwhelmed by the dual feedback from teachers and AI, especially when the suggestions were contradictory. Participants explained that they needed time to decide which feedback to prioritize. The qualitative data reflects this challenge. One respondent said, “Sometimes, the volume of feedback from both sources was overwhelming, to the point of confusion about which suggestions to follow,” while another noted, “The AI feedback was sometimes overly formal, so I still had to select which suggestions were appropriate for the context of the assignment.” Over time, however, students reported becoming more selective and critical, which ultimately strengthened their judgment.

4.3.5. Perceived improvement in writing quality

The students agreed that hybrid feedback significantly improved the overall quality of their writing. They noted considerable improvements in clarity, coherence, grammatical correctness, and argumentation structure. This sentiment was also reflected in the qualitative feedback. One participant reflected, “Initially, my paragraphs were disorganized, but after receiving feedback from both the teacher and the AI, the flow became much clearer,” Another participant explained, “My revisions became more academic because the AI refined the language while the teacher improved the content.” This perception is consistent with the quantitative results, which show an improvement in scores across all components of the rubric.

The convergence of quantitative and qualitative data shows that students generally have a positive perception of the integration of feedback from teachers and feedback from artificial intelligence (AI) in the context of academic writing. The survey results show moderate to high agreement on all items, with average scores ranging from 3.16 to 3.52, suggesting that students view hybrid feedback as beneficial for improving accuracy, clarity, and overall writing skills. Items with higher average scores indicate that students place a high value on the clarity and usefulness of combined feedback, while items with moderate scores reflect the cognitive load associated with processing input from two sources. The data from the qualitative interviews support these findings by highlighting the complementary nature of hybrid feedback: AI provides immediate and detailed linguistic corrections, while teacher feedback promotes the development of ideas, arguments, and contextual appropriateness. Furthermore, participants report increased confidence, greater awareness of writing weaknesses, and more intensive involvement in the revision process when both sources of feedback are available. Overall, these findings show that participants view the integrated feedback from teachers and AI as a comprehensive support mechanism that improves both linguistic accuracy and rhetorical

development, thereby increasing the effectiveness of revision practices and overall writing performance.

5. Discussion

This study demonstrates that hybrid feedback creates a scaffolding effect where AI addresses lower-order concerns, freeing cognitive resources for teacher feedback on higher-order concerns. As a result, the significant increase in total writing scores suggests that the hybrid feedback model provides a more comprehensive and effective form of support compared to using a single feedback source alone. Both descriptive and inferential analyses show consistent progress after the intervention, indicating that the use of a combination of human expertise and automated feedback mechanisms helps improve language accuracy as well as conceptual depth. These results reaffirm previous research emphasizing the increasingly important role of AI-supported feedback in writing pedagogy (Dizon & Gold, 2023; Escalante et al., 2023; Guan et al., 2024; Roa & Halim, 2024; Zainuddin, 2004).

The substantial effect size observed in this study ($d = 1.78$), far exceeding the conventional effects reported in previous meta-analyses on Automated Writing Evaluation (AWE), which typically range from 0.4 to 0.6. These findings confirm the superiority of the hybrid approach over exclusive reliance on AI, while also demonstrating that the improvements are not only statistically significant but also educationally meaningful. These results support the theoretical framework underlying the hybrid feedback model, according to which teacher feedback provides interpretative and contextual insights, while AI-based systems deliver quick and precise linguistic corrections (Cheng & Zhang 2021; Ranalli 2021; Wilson et al., 2024). From a socio-cognitive perspective, writing development improves when learners receive iterative and multidimensional input. This study expands this perspective by showing that such multi-layered support mechanisms can be effectively applied in an AI-mediated instructional context.

The improvements observed in all five writing components provide a more detailed understanding of the operational dynamics of hybrid feedback. Enhancements in content and organization indicate that teacher feedback plays a crucial role in strengthening idea development and rhetorical coherence. Previous research has repeatedly emphasized the irreplaceable role of human feedback in addressing high-level writing issues that require contextual and interpretative judgment (Cheng & Zhang 2021; Lo et al. 2025). Conversely, the progress observed in vocabulary, language use, and mechanics highlights the capability of AI-generated feedback to detect grammatical inaccuracies and improve sentence-level accuracy (Dizon & Gayed 2024; Fleckenstein et al., 2023). The integration of these respective strengths underscores the complementary potential of human expertise and artificial intelligence in the didactics of writing instruction.

The perception of the students increasingly enhances the effectiveness of the hybrid approach applied in this study. The survey results show that, in general, there is a positive attitude towards the clarity and usefulness of the integrated feedback, while the interview results indicate that students view feedback from instructors and AI as two components that have different functions but complement each other. This perception is consistent with previous

studies, which state that automatic feedback can increase learners' engagement and willingness to revise their texts, especially when combined with instructor guidance (Huang et al., 2024; Song & Song 2023). In addition, the students also reported an increased awareness of weaknesses in their texts as well as increased confidence in revising drafts, suggesting that this intervention not only supports the improvement of technical aspects of writing but also strengthens the students' ability to take action in their learning.

Although students generally show a positive attitude, some experience initial cognitive load when receiving feedback from two different sources. This observation is consistent with concerns highlighted in previous studies, which indicate that AI recommendations may require careful interpretation and critical evaluation (Dergaa et al., 2023; Dizon & Gayed, 2024). Nevertheless, the more students get accustomed to this process, the more they develop a selective assessment of which feedback should be prioritized. This suggests that hybrid feedback is most effective when teachers actively support students in interpreting the AI-generated suggestions.

Overall, these results contribute to the ongoing dialogue about the role of AI in education. Rather than functioning as a replacement for teachers, AI appears to be most effective as a complementary tool that enhances both teaching efficiency and linguistic accuracy (Escalante et al., 2023; Fleckenstein et al., 2023). The consistently observed improvements across all rubric components, along with significant effect sizes, provide strong empirical support for the integration of hybrid feedback models in teaching academic writing in English as a foreign language (EFL). These findings suggest that a combined approach of teacher and AI feedback can create a valuable framework to promote both conceptual clarity and linguistic accuracy among developing writers.

6. Conclusion

This study provides compelling evidence that the integration of teacher feedback and AI-generated responses leads to meaningful and multidimensional enhancement of EFL students' academic writing. By combining human evaluative insight with automated linguistic analysis, learners received both conceptual direction and precise, form-focused corrections. This dual-support system resulted in significant improvement not only in overall writing performance but also across each analytic component of the assessment rubric. Furthermore, students' positive evaluations and reflective comments indicate that hybrid feedback contributes not only to greater textual accuracy and coherence but also to increased confidence, heightened metacognitive awareness, and deeper engagement in the revision process.

The findings add to the expanding scholarship on hybrid intelligence in writing pedagogy by demonstrating that teacher and AI feedback generate the greatest benefit when implemented in a complementary rather than isolated manner. Moreover, the study offers empirical validation for instructional models that purposefully integrate digital technologies with sustained teacher mediation. Such models appear capable of addressing higher-order concerns (e.g., argumentation and organization) alongside lower-order issues (e.g., grammar and mechanics) in a balanced and pedagogically efficient way. Taken together, this research provides a strong evidence-based rationale for adopting hybrid feedback systems in academic writing courses and

highlights their potential as a forward-looking approach within increasingly AI-mediated EFL learning environments. Nevertheless, hybrid feedback is not a panacea. Its successful implementation requires digital literacy training for students and clear institutional guidelines to prevent over-reliance on AI-generated corrections.

Declaration on the use of AI

The authors used AI applications, including Grammarly and Quillbot, to enhance the quality and structure of the English language, making the text more readable and comprehensible. The ideas and content presented in this manuscript are the original work of the authors themselves.

References

- Baroudi, S., Mubeen, N., & Karaki, S. (2025). An action research study on AI video vs. written feedback: enhancing undergraduate academic writing and critical thinking. *Cogent Education*, 12(1), 1-19. <https://doi.org/10.1080/2331186X.2025.2527302>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745.
- Cheng, H.-F., & Dörnyei, Z. (2007). The use of motivational strategies in language instruction: The case of EFL teaching in Taiwan. *Innovation in Language Learning and Teaching*, 1(1), 153–174. <https://doi.org/10.2167/illt048.0>
- Cheng, X., & Zhang, L. J. (2021). Sustaining university English as a foreign language learners' writing performance through provision of comprehensive written corrective feedback. *Sustainability (Switzerland)*, 13(15). <https://doi.org/10.3390/su13158192>
- Coffin, C., Curry, M. J., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2005). *Teaching academic writing: A toolkit for higher education*. Routledge.
- Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. Ben. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, 40(2), 615–622. <https://doi.org/10.5114/BIOLSPORT.2023.125623>
- Dizon, G., & Gayed, J. M. (2024). A systematic review of Grammarly in L2 English writing contexts. *Cogent Education*, 11(1), 2397882. <https://doi.org/10.1080/2331186X.2024.2397882>
- Dizon, G., & Gold, J. (2023). Exploring the effects of Grammarly on EFL students' foreign language anxiety and learner autonomy. *JALT CALL Journal*, 19(3), 299–316. <https://doi.org/10.29140/jaltcall.v19n3.1049>
- Do, H. M., & Phan, H. L. T. (n.d.). *Teacher written feedback on higher order skills of writing: A case study of L2 Vietnamese students*.
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1). <https://doi.org/10.1186/s41239-023-00425-2>
- Ferris, D. (2011). *Treatment of error in second language student writing*. University of Michigan Press.

- Ferris, D. R. (2014). Responding to student writing: Teachers' philosophies and practices. *Assessing Writing*, 19, 6–23.
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: a multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.1162454>
- Guan, L., Li, S., & Gu, M. M. (2024). AI in informal digital English learning: A meta-analysis of its effectiveness on proficiency, motivation, and self-regulation. *Computers and Education: Artificial Intelligence*, 7(October), 100323. <https://doi.org/10.1016/j.caeai.2024.100323>
- Huang, X., Xu, W., Li, F., & Yu, Z. (2024). A meta-analysis of effects of automated writing evaluation on anxiety, motivation, and second language writing skills. *Asia-Pacific Education Researcher*, 33(4), 957–976. <https://doi.org/10.1007/s40299-024-00865-y>
- Hyland, K. (2019). *Second language writing*. Cambridge university press.
- Hyland, K. (2020). *English for specific purposes: what is this and where is it taking us?* 10, 1–23.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, 39(2), 83–101.
- Jacobs, H. L. (1981). *Testing ESL composition: A practical approach*. English composition program. ERIC.
- Koltovskaia, S. (2023). Postsecondary L2 writing teachers' use and perceptions of Grammarly as a complement to their feedback. *ReCALL*, 35(3), 290–304.
- Laksmi, E. D. (2006). Scaffolding students' writing in EFL class: Implementing process approach. *Teflin Journal*, 17(2), 144–156.
- 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), 1–16. <https://doi.org/10.1177/21582440251352907>
- Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence*, 3, 100080.
- Mendoza Anchundia, K. I., Espín Bravo, E. P. E., Sánchez Zavala, D. A., Galeas Galeas, E. V., & Mina Acosta, J. Y. (2025). Estado del arte sobre la escritura académica en universitarios: competencias, falencias y enfoques de mejora. *Boletín Científico Ideas y Voces*, 5(3), 251–264. <https://doi.org/10.60100/bciv.v5i3.123>
- Mishra, K. C. (2023). *ESL Academic Writing: A course curriculum necessity for the higher education institutes of Arunachal Pradesh*.
- 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.
- Ramzan, M., Mushtaq, A., & Ashraf, Z. (2023). Evacuation of difficulties and challenges for academic writing in ESL learning. *The University of Chitral Journal of Linguistics and Literature (JLL)*, 7(1), 42–49.
- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52, 100816. <https://doi.org/https://doi.org/10.1016/j.jslw.2021.100816>

- Roa, A. A. P., & Halim, S. (2024). The impact of AI-powered software on second language (L2) writing: A systematic literature review. *Research and Innovation in Applied Linguistics-Electronic Journal*, 2(2), 138–158.
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14(December), 1–14. <https://doi.org/10.3389/fpsyg.2023.1260843>
- Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2024.101894>
- Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19, 51–65.
- Swales, J.M.; Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). University of Michigan Press.
- Weber, F., Wambsganss, T., & Söllner, M. (2025). Enhancing legal writing skills: The impact of formative feedback in a hybrid intelligence learning environment. *British Journal of Educational Technology*, 56(2), 650–677. <https://doi.org/10.1111/bjet.13529>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Wilson, J., Palermo, C., & Wibowo, A. (2024). Elementary English learners' engagement with automated feedback. *Learning and Instruction*, 91, 101890. <https://doi.org/https://doi.org/10.1016/j.learninstruc.2024.101890>
- Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58(1), 87–125.
- Yang, L. (Francoise), Zhang, L. J., & Dixon, H. R. (2023). Understanding the impact of teacher feedback on EFL students' use of self-regulated writing strategies. *Journal of Second Language Writing*, 60, 101015. <https://doi.org/10.1016/j.jslw.2023.101015>
- Zainuddin, M. (2004). The effect of giving feedback to students' writing. *TEFLIN Journal*, 15(2), 117–126.