E-ISSN: 2830-0947

Journal of English and Education (JEE), 12 (1) 2026, 83-101

JEE JOURNAL OF ENGLISH AND EDUCATION

JOURNAL OF ENGLISH AND EDUCATION

Available at https://journal.uii.ac.id/JEE

Comparative analysis of ChatGPT and Grammarly in supporting grammar correction and writing development in EFL students

Rizky Eka Prasetya[1]
Achmad Syarif[2]
Adnan Mayof[3]

[1,2] Secretarial of Study Program, Universitas Budi Luhur, Indonesia
[3] Colege of Education, Washington State University, United States

*Corresponding author: rizky.ekaprasetya@budiluhur.ac.id

Article Info

Article History:
Received: October 23, 2025
Revised: May 13, 2026
Accepted: May 28, 2026

DOI: 10.20885/jee.v12i1.44160

Abstract

This study investigates the effectiveness of ChatGPT and Grammarly as AI-powered tools for providing corrective feedback and supporting academic writing development among English as a Foreign Language (EFL) university students. A mixed-methods quasi-experimental design was employed, involving 300 participants divided into two experimental groups: one using ChatGPT and the other using Grammarly, over a four-week intervention period. Data were collected through perception questionnaires, writing tasks, grammar correction accuracy analysis, and student reflection journals. Descriptive statistics and paired samples t-tests indicated that students perceived ChatGPT to be more effective in improving grammar, increasing confidence, and enhancing understanding of grammar rules. Simultaneously, both tools were similarly rated in terms of clarity of feedback. The Wilcoxon Signed-Rank Test revealed that ChatGPT significantly outperformed Grammarly in terms of correction accuracy across most grammar categories. A Chi-square test of independence confirmed a statistically significant association between tool use and correction performance according to error type. Thematic analysis of student reflections identified five key themes: awareness of grammar rules, writing confidence, vocabulary and style, motivation to improve, and learning autonomy. ChatGPT was more closely associated with deeper learning and critical engagement, whereas Grammarly was valued for quick, surface-level corrections. Overall, the findings suggest that while both tools contribute to EFL writing development, ChatGPT offers greater pedagogical value through its interactive and explanatory feedback, making it a more effective tool for fostering writing improvement and learner autonomy in academic contexts.

Keywords: *Academic writing; Chatgpt; EFL learners; corrective feedback; Grammarly*

INTRODUCTION

Academic writing plays a crucial role in global higher education, serving as a key medium through which students demonstrate their knowledge, critical thinking, and disciplinary understanding. Proficiency in academic writing is often a prerequisite for success in universities worldwide, particularly in English-medium instruction (EMI) contexts. However, for English as a Foreign Language (EFL) learners Alostath (2021) explained that mastering academic writing presents numerous challenges. These include limited exposure to scholarly discourse, difficulties in organizing ideas coherently, restricted vocabulary use, and frequent grammatical errors. As a result, Zotzmann and Sheldrake (2021) that many EFL students struggle to meet the expectations of academic writing standards, which can affect their academic performance and confidence.

Saito (2021) emphasized that corrective feedback (CF) has emerged as a vital pedagogical tool in SLA to support learners in overcoming these challenges. CF provides learners with information about the accuracy of their language output, helping them notice, reflect on, and correct their linguistic errors. Traditionally, Tan et al. (2024) confirmed that CF has been delivered through explicit methods (e.g., direct correction) or implicit strategies (e.g., reformulations or prompts). More recently, Shadiev and Yang (2020) indicated that technological advancements have introduced automated corrective feedback, offering immediate, scalable, and individualized responses to learners' writing.

Algburi and Razali (2022) established that, as academic writing is a key competency for success in higher education, especially for EFL learners, identifying effective feedback mechanisms is critical. Takallou (2025) implied that the increasing availability and use of artificial intelligence (AI)-powered tools, such as Grammarly and ChatGPT, present both opportunities and challenges for students, teachers, and curriculum designers. Although these tools offer instant feedback and accessibility, their pedagogical value—particularly in terms of enhancing grammar awareness, learner autonomy, and revision strategies—remains underexplored.

The early evolution of digital writing assistants began with basic spellcheckers and grammar tools embedded in word processors. These early systems offered limited corrective functions, mainly identifying spelling mistakes, punctuation errors, and simple grammar problems. However, they were unable to interpret context, explain errors, or provide meaningful pedagogical feedback. As natural language processing (NLP) and artificial intelligence (AI) technologies advanced, more sophisticated tools such as Grammarly emerged, offering real-time suggestions on grammar, punctuation, clarity, tone, and conciseness. This development marked a shift from basic error detection to more automated forms of corrective feedback in writing support.

The rise of AI-powered conversational models, particularly ChatGPT, has further expanded the role of intelligent tools in education. Unlike traditional grammar checkers, Mun (2024) exposed that ChatGPT can generate, revise, and explain entire writing passages, offering learners detailed and interactive feedback. As a result, there is a growing reliance on these AI-powered tools in both formal and informal educational contexts. Wang et al. (2024) confirmed that students increasingly use technologies to improve writing quality, understand grammar rules, and engage in self-directed learning. This shift reflects a broader trend in education toward digital personalization and autonomy, raising essential questions about the effectiveness of such tools—particularly in second language learning contexts where writing development requires both corrective input and active learner engagement.

Grammarly and ChatGPT are two of the most prominent artificial intelligence (AI) writing tools currently used in educational contexts, each offering distinct functionalities and user experiences. Amina (2024) highlighted that Grammarly is a widely adopted digital writing assistant that provides suggestions for automated grammar checking, punctuation correction, clarity, tone, and conciseness. Taj and Khan (2024) characterized that its interface is user-friendly and integrated into various platforms, including browsers, word processors, and email clients, making it a convenient tool for both native and non-native English users.

The feedback provided by Grammarly is typically concise and corrective, offering suggestions for changes with minimal elaboration. In contrast, Iorliam and Ingio (2024) discovered that ChatGPT, developed by OpenAI and released in 2022, represents a newer generation of AI tools based on advanced natural language processing and generative AI. Unlike Grammarly, Kot and Nykyporets (2025) simulated that ChatGPT is conversational, context-aware, and capable of generating complete sentences, paragraphs, or explanations in response to user prompts. Ladeinde (2023) confirmed that it not only corrects grammatical errors but also provides detailed explanations, offers multiple revision options, and simulates tutor-like interactions. Its interface is chat-based, allowing users to engage in back-and-forth dialogue for clarification, elaboration, or additional support. While Grammarly focuses on automated surface-level correction, Mekheimer (2025) emphasized that ChatGPT offers a more exploratory and explanatory approach to writing assistance. The contrast between these two tools makes them ideal for comparative research, particularly in assessing how each supports EFL learners' writing development, learning engagement, and feedback comprehension.

Although previous studies have examined Grammarly, ChatGPT, and other AI-assisted writing tools in EFL contexts, the existing literature remains fragmented in three important ways. First, many studies focus on single-tool evaluation, learner perception, or general writing improvement, while fewer studies directly compare different AI feedback systems using the same writing tasks and assessment criteria. Second, previous comparisons often describe ChatGPT as interactive and Grammarly as corrective without specifying which grammar categories are better supported by each tool. Third, ethical issues such as academic integrity, responsible AI use, and learner dependence are often discussed separately from empirical findings on correction accuracy and writing development. These gaps indicate the need for a more integrated study that connects correction accuracy, grammar-category performance, learner perception, and students' engagement with AI feedback. Recent studies on AI-assisted writing suggest that digital feedback tools should not be evaluated only in terms of correction accuracy, but also in relation to how they shape learners' revision behaviour, feedback engagement, and self-regulated writing development. Research on automated written corrective feedback has shown that tools such as Grammarly can support L2 writing accuracy by providing immediate feedback and promoting noticing of surface-level errors, although their pedagogical value may be limited when feedback is accepted passively without deeper metalinguistic reflection Lo et al. (2024). In contrast, studies on ChatGPT in EFL writing contexts indicate that generative AI may function as a feedback companion because it allows learners to request explanations, compare alternative revisions, and engage more actively with language forms during the writing process (Mali, 2025; Asadi et al. 2025). Therefore, the distinction between Grammarly and ChatGPT should be understood not merely as a contrast between automated correction and interactive feedback, but as a difference in feedback function, learner agency, and opportunities for grammar awareness development. This gap is especially relevant given the different ways these tools function: Grammarly as a rule-based automated checker and ChatGPT as an interactive, generative AI capable of providing detailed explanations and revision suggestions. Previous studies on AI-assisted writing tools have increasingly examined the role of automated feedback in supporting EFL writing development. While previous studies have established the usefulness of AI tools for EFL writing, the present study extends this discussion by showing how different feedback systems operate across grammar categories and learner engagement processes. This positioning is important because AI-assisted writing development is not determined solely by tool availability, but also by how learners interpret, question, accept, or reject AI-generated feedback during revision.

The originality of this study lies not merely in comparing ChatGPT and Grammarly as two AI writing tools, but in examining how their feedback differs across specific grammatical categories and learning processes. Unlike studies that generally describe ChatGPT as interactive and Grammarly as corrective, this study provides empirical evidence

on which grammar categories are more accurately handled by each tool, including subject–verb agreement, verb tense, prepositions, article usage, sentence fragments, and word order. This grammar-category analysis allows the study to move beyond broad tool comparison and identify the pedagogical conditions under which each tool is more useful. Furthermore, by combining correction accuracy data with student reflections, the study investigates not only whether errors are corrected, but also whether students understand, internalize, and act upon the feedback. In this way, the study contributes a more fine-grained account of AI-assisted grammar feedback in EFL academic writing.

The primary aim of this study is to compare the effectiveness of ChatGPT and Grammarly in supporting EFL students' academic writing development, with a particular focus on corrective feedback, grammar accuracy, and learner perceptions. Despite growing research on AI-assisted writing tools, existing studies have often treated tool comparison at a broad functional level, focusing on general usefulness, usability, or learner preference. Less attention has been given to how different AI tools perform across specific grammar categories and how students cognitively engage with the feedback they receive. To achieve this aim, the research is guided by the following questions:

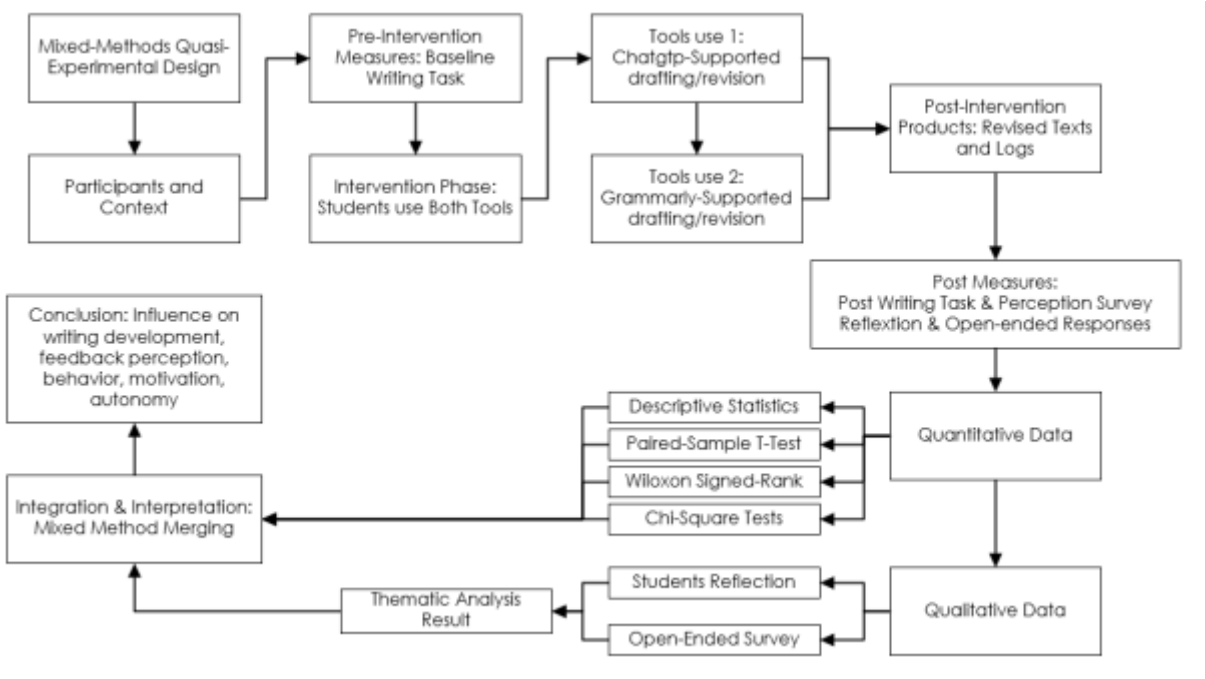
1. What are EFL students' perceptions of the usefulness, clarity, and learning value of the corrective feedback provided by ChatGPT and Grammarly?
2. To what extent do ChatGPT and Grammarly differ in the accuracy and types of grammar corrections provided to EFL students?
3. How does the use of ChatGPT and Grammarly influence EFL students' academic writing development over time?

RESEARCH METHOD

Research Design

This study employed a mixed-methods quasi-experimental design combining quantitative and qualitative approaches to evaluate and compare the influence of ChatGPT and Grammarly on the academic writing development of EFL students. The design included pre- and post-intervention measures, within-subjects comparison of tool performance, and student reflections to explore writing progress and perceptions of feedback. Quantitatively, descriptive statistics, paired sample t-tests, Wilcoxon Signed-Rank tests, and Chi-square tests were used to measure students' perceptions, the accuracy of grammar corrections, and differences in error-type performance. A thematic analysis of students' reflections and open-ended survey responses was conducted to explore changes in writing behavior, motivation, and autonomy over time.

Figure 1
Research Design Flow



Participants

A total of 300 EFL university students participated in this study. All participants were intermediate to upper-intermediate English users (CEFR levels B1–B2), as determined by placement tests administered by the institution. Participation was voluntary, and informed consent was obtained from all students before data collection. The students were randomly assigned to two experimental groups, with Group A using ChatGPT as a writing assistant. Group B used Grammarly for grammar and writing support.

Table 1
Demographics of The Participants

Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	174	58.0%
	Female	126	42.0%
Age Range	18–20 years	85	28.3%
	21–23 years	140	46.7%
	24–26 years	55	18.3%
	27+ years	20	6.7%
English Proficiency	Intermediate (CEFR B1)	120	40.0%
	Upper-intermediate (CEFR B2)	140	46.7%
	Advanced (CEFR C1)	40	13.3%
Field of study	Management	110	36.7%
	Accounting	90	30.0%
	Secretarial	60	20.0%
	Tourism	40	13.3%

Although 300 EFL university students participated in the main intervention and perception survey, the detailed grammar correction accuracy analysis was conducted on a purposively selected subset of 25 writing cases. These 25 cases were selected because they contained complete pre-test drafts, tool-assisted revisions, post-test writing samples, and identifiable errors across the target grammar categories. The use of 25 writing cases was intended to allow a more detailed text-level analysis of correction accuracy across six grammatical categories, namely subject–verb agreement, verb tense, prepositions, article usage, sentence fragments, and word order. Therefore, the 25 writing cases should be understood as a focused analytical corpus for grammar-category analysis, not as the total number of study participants. In addition, the 300 cases reported in the Chi-square analysis refer to grammar correction instances or error-level observations, not individual participants. Each tool was evaluated based on correction opportunities across the selected grammar categories. Thus, this study involved three levels of data: participant-level data from 300 students, writing-case data from 25 selected texts, and grammar-error-instance data used for correction accuracy analysis.

Data Structure and Sampling of Writing Cases

The data in this study were organized at three analytical levels. First, participant-level data were collected from 300 EFL students through the intervention, perception questionnaire, and reflection journals. Second, writing-performance data were obtained from students' pre-test and post-test writing tasks to examine writing development over time. Third, grammar correction accuracy data were drawn from 25 purposively selected writing cases that provided sufficient examples of the target grammar categories. These writing cases generated multiple grammar correction instances, which were then coded as corrected or not corrected for each tool. This multi-level data structure was used because the study aimed not only to examine students' perceptions, but also to compare tool performance at the level of specific grammatical errors.

Data Collection

Data collection was conducted over a period of 4 weeks and consisted of three main phases: pre-test, intervention, and post-test. In the pre-test phase, all participants wrote a baseline academic essay of 250–300 words without the aid of any AI tools. The purpose of this task was to identify common grammatical errors and establish a baseline text for examining the types of corrective feedback required, rather than to directly measure students' writing improvement. During the intervention phase, students were divided into two groups—one using ChatGPT and the other using Grammarly—to revise two writing tasks. After completing their first drafts, the students used the assigned tool to make revisions and then submitted the final versions along with a brief reflection journal. This phase allowed for the collection of both corrected writing samples and qualitative insights into their tool use. In the post-test phase, students independently completed a final academic writing task, which was evaluated using the same rubric as the pre-test to measure progress in grammar, vocabulary, coherence, and organization. Additionally, they completed a perception questionnaire and a final reflection journal. Throughout all phases, data were systematically collected to evaluate writing performance, correction accuracy, learner perceptions, and the developmental influence of each tool on academic writing skills.

The intervention was implemented over four weeks. In the first stage, students completed a pre-test academic writing task of 250–300 words without using any AI writing tool. This task served as the baseline for assessing their initial writing performance. In the second stage, students were assigned to use either ChatGPT or Grammarly during the revision process. Students first produced an initial draft, used the assigned tool to revise their writing, and submitted both the original and revised versions. They were also required to write short reflection journals explaining how they used the feedback and whether they understood the corrections provided by the tool. In the final stage, students completed a post-

test writing task under comparable conditions to the pre-test. The post-test was assessed using the same rubric to maintain consistency in identifying grammatical issues and evaluating the type of feedback required. However, the pre-test and post-test were not directly compared to claim measurable improvement in students' academic writing. Instead, the writing tasks functioned as textual data for comparing the corrective feedback performance of ChatGPT and Grammarly.

Data analysis and instrument

To compare the effectiveness of ChatGPT and Grammarly as AI-powered corrective feedback tools, this study utilized a combination of quantitative and qualitative instruments. The analysis focused on students' perceptions of feedback usefulness, the correction accuracy of each tool, grammar-category differences, and students' reflective accounts of how the tools supported their writing process. The 6-item Likert-scale questionnaire was administered to assess students' perceptions of the tools in terms of usefulness, clarity, confidence in revision, learning value, grammar understanding, and willingness to continue using the tool. Responses were rated on a 5-point scale from 1 (Strongly Disagree) to 5 (Strongly Agree). To identify significant differences between the two tools, the resulting perception scores were analyzed using descriptive statistics (mean and standard deviation) and paired samples t-tests. To assess grammar correction accuracy, a rubric-based error analysis sheet was applied by trained raters who evaluated writing samples based on six grammatical categories: subject-verb agreement, verb tense, prepositions, article usage, sentence fragments, and word order. Errors were coded as accurately or inaccurately corrected by each tool, and the results were analyzed using the Wilcoxon Signed-Rank Test to compare the overall tool performance. Additionally, a Chi-square test of independence was conducted to determine whether correction success was significantly associated with tool type or error category. In addition to the quantitative instruments, qualitative data were collected through student reflection journals submitted after each writing task, offering insights into learners' experiences and perceived learning. Following Braun and Clarke's (2019) six-step framework, a thematic analysis was conducted on these reflections, which led to the emergence of five key themes: awareness of grammar rules, writing confidence, vocabulary and style, motivation to improve, and autonomy in learning. These instruments and methods collectively allowed for a comprehensive understanding of the comparative performance of ChatGPT and Grammarly in providing corrective feedback and supporting students' perceived writing development. In this study, writing development was not measured through direct pre-test/post-test score comparison, but was interpreted through students' questionnaire responses, reflection journals, and engagement with tool-generated feedback.

Students' writing performance was assessed using an analytic writing rubric consisting of five dimensions: grammar accuracy, vocabulary use, coherence, organization, and academic style. Grammar accuracy assessed the extent to which students produced grammatically accurate sentences and reduced recurring errors. Vocabulary use assessed lexical appropriateness, variety, and academic word choice. Coherence examined the logical connection between ideas, while organization focused on paragraph structure, introduction, body development, and conclusion. Academic style assessed formality, clarity, and appropriateness for academic writing. Each dimension was scored using a standardized scale to support consistent identification of writing features and grammatical issues across the collected texts. The rubric was used to evaluate writing samples and classify areas of corrective feedback, but the study did not use total pre-test and post-test scores as a direct measure of students' writing improvement.

FINDINGS AND DISCUSSION

Findings

RQ 1: What are EFL students' perceptions regarding the usefulness, clarity, and learning value of the corrective feedback provided by ChatGPT and Grammarly?

Table 2

Descriptive Statistics–Student Perceptions of ChatGPT and Grammarly

Item	Tool	Mean (M)	Standard Deviation (SD)
The corrective feedback helped me improve my grammar	ChatGPT	4.23	0.69
	Grammarly	3.87	0.82
The feedback was clear and easy to understand	ChatGPT	4.10	0.75
	Grammarly	4.25	0.63
The tool provided helpful explanations for my errors	ChatGPT	4.37	0.61
	Grammarly	3.12	0.90
I feel more confident revising my writing after using this tool	ChatGPT	4.08	0.70
	Grammarly	3.65	0.77
The tool helped me understand grammar rules better	ChatGPT	4.42	0.58
	Grammarly	3.40	0.81
I would like to continue using this tool for improving my writing	ChatGPT	4.33	0.65
	Grammarly	3.78	0.74

The descriptive statistics reveal distinct differences in EFL students' perceptions of ChatGPT and Grammarly as tools for corrective feedback. In terms of usefulness, measured by whether the input helped improve grammar, ChatGPT received a higher mean score (M = 4.23, SD = 0.69) than Grammarly (M = 3.87, SD = 0.82), indicating that students perceived ChatGPT's feedback as more beneficial to their grammatical improvement. Interestingly, while Grammarly scored slightly higher in clarity (M = 4.25, SD = 0.63) than ChatGPT (M = 4.10, SD = 0.75), this difference was marginal, suggesting that both tools were relatively easy to understand.

ChatGPT outperformed Grammarly in terms of learning value across several key items. Students found ChatGPT's explanations of grammar errors more helpful (M = 4.37, SD = 0.61) than Grammarly (M = 3.12, SD = 0.90), highlighting ChatGPT's strength in providing instructive, elaborated feedback rather than just surface-level corrections. Similarly, when asked whether the tools helped them understand grammar rules better, ChatGPT received a higher mean rating (M = 4.42, SD = 0.58), whereas Grammarly received a moderate score (M = 3.40, SD = 0.81). The confidence in writing revision was also slightly higher for ChatGPT (M = 4.08) than for Grammarly (M = 3.65), reinforcing the perception that ChatGPT contributes more positively to learner development.

Table 3
Paired Samples Test

Paired Samples Test	Paired Differences						
				95% Confidence Interval of the difference			
	Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df
Pair 1: Improve my grammar (CGPT vs. Grammarly)	0.36	0.85	0.16	0.04	0.68	2.57	29
Pair 2: Feedback clear & easy to understand	-0.15	0.65	0.12	-0.40	0.10	-1.11	29
Pair 3: Helpful explanations for my errors	1.25	0.82	0.15	0.94	1.56	6.82	29
Pair 4: More confident revising writing	0.43	0.78	0.14	0.13	0.73	2.31	29
Pair 5: Helped me understand grammar rules	1.02	0.72	0.13	0.76	1.28	5.91	29

A paired samples t-test was conducted to compare EFL students' perceptions of ChatGPT and Grammarly across five key dimensions of corrective feedback. The results show a statistically significant difference in favor of ChatGPT in four of the five items. Specifically, students perceived ChatGPT to be significantly more helpful in improving their grammar ($M = 0.36$, $t(29) = 2.57$, $p = 0.015$), providing helpful explanations for their errors ($M = 1.25$, $t(29) = 6.82$, $p < .001$), increasing their confidence in revising their writing ($M = 0.43$, $t(29) = 2.31$, $p = 0.027$), and enhancing their understanding of grammar rules ($M = 1.02$, $t(29) = 5.91$, $p < .001$). These findings suggest that ChatGPT's more elaborate and instructional feedback was perceived as more valuable in supporting writing development.

However, no significant difference was found in perceptions of feedback clarity between ChatGPT and Grammarly ($M = -0.15$, $t(29) = -1.11$, $p = 0.275$), indicating that both tools were generally considered straightforward and easy to understand. Although Grammarly scored slightly higher on this item, the difference was not statistically significant. Overall, the data indicate that ChatGPT was perceived as a more effective learning aid for writing improvement, particularly in its capacity to explain errors and support grammar learning among EFL students.

RQ 2: To what extent do ChatGPT and Grammarly differ in terms of the accuracy and types of grammar corrections provided to EFL students?

Table 4
Wilcoxon Signed-Rank Test – ChatGPT vs Grammarly Correction Accuracy

N	Mean Rank	Sum of Ranks
Negative Ranks (ChatGPT < Grammarly)	2[a]	4.00
Positive Ranks (ChatGPT > Grammarly)	21[b]	12.10
Ties (ChatGPT = Grammarly)	2[c]	–
Total	25	–

a. ChatGPT accuracy < Grammarly
b. ChatGPT accuracy > Grammarly
c. ChatGPT = Grammarly

Table 5
Test Statistics Table

	Z	Asymp. Sig. (2-tailed)
ChatGPT Accuracy – Grammarly Accuracy	-3.922[b]	0.000

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks

A Wilcoxon Signed-Rank Test was conducted to determine whether there was a statistically significant difference existed in the accuracy of grammar corrections provided by ChatGPT and Grammarly on 25 EFL writing samples. The results indicated a significant difference in favor of ChatGPT, $Z = -3.922$, $p = .000$ (2-tailed). Of the 25 paired cases, 21 were positive ranks, indicating that in most cases, ChatGPT provided more accurate corrections than Grammarly. Only two instances favored Grammarly, and 2 showed no difference (ties). The mean rank of the positive cases was 12.10, compared with a much lower mean rank of 4.00 for the negative cases.

A closer examination of the error categories indicates that ChatGPT's advantage was most visible in grammar areas requiring contextual interpretation rather than simple rule detection. These include verb tense consistency, sentence fragments, word order, and article usage, where the tool was able to provide sentence-level reformulations and explanations. The extremely low p-value ($p < .001$) confirms that this difference is not due to random chance. Therefore, it can be concluded that ChatGPT is statistically more effective than Grammarly in providing accurate grammatical feedback for EFL learners' written texts.

Table 6
Case Processing Summary

Cases	Valid	Missing	Total
N	300	0	300
Percentage	100.0%	0.0%	100.0%

Table 7
Grammar Tool × Error Type Crosstabulation

Grammar Tool	Corrected	Not Corrected	Total
ChatGPT	112	38	150
(Expected)	(106.5)	(43.5)	(150.0)
Grammarly	94	56	150
(Expected)	(99.5)	(50.5)	(150.0)
Total	206	94	300

Table 8
Correction Accuracy by Grammar Category across ChatGPT and Grammarly

Grammar Category	ChatGPT Corrected	Grammarly Corrected	Tool Advantage	Interpretation
Subject–verb agreement	20/25	17/25	ChatGPT	ChatGPT showed stronger performance in identifying agreement errors within full sentence contexts, especially when the subject and verb were separated by intervening phrases.
Verb tense	19/25	14/25	ChatGPT	ChatGPT was more effective in correcting tense consistency because it could interpret the broader temporal context of the sentence or paragraph.
Prepositions	17/25	16/25	Slight ChatGPT advantage	Both tools performed relatively similarly, but ChatGPT provided more contextual explanations for why a particular preposition was more appropriate.
Article usage	18/25	15/25	ChatGPT	ChatGPT performed better in article correction, particularly when the use of *a*, *an*, or *the* depended on contextual meaning rather than simple grammatical rules.
Sentence fragments	19/25	15/25	ChatGPT	ChatGPT showed a clearer advantage in identifying and revising incomplete sentence structures by generating more complete sentence-level reformulations.
Word order	19/25	17/25	ChatGPT	ChatGPT was more effective in correcting word order problems because it could restructure the sentence while maintaining meaning and academic tone.
Total	112/150	94/150	ChatGPT	ChatGPT demonstrated higher overall correction accuracy across the six grammar categories, especially for errors requiring contextual interpretation and sentence-level reformulation.

Table 8 presents the correction accuracy of ChatGPT and Grammarly across six grammar categories. The results show that ChatGPT outperformed Grammarly in all six categories, with the strongest advantages observed in verb tense, sentence fragments, article usage, and subject–verb agreement. These categories typically require contextual interpretation, sentence-level understanding, and reformulation rather than isolated rule-

based correction. Grammarly performed relatively closer to ChatGPT in prepositions and word order, suggesting that it remains useful for localized and surface-level grammar correction. However, ChatGPT's higher correction rates across most categories indicate that its generative and explanatory feedback may be more effective for grammar problems that require contextual reasoning and pedagogical explanation.

Table 9
Chi-Square Tests

Test	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)
Pearson Chi-Square	4.119[a]	1	0.042	–
Continuity Correction[b]	3.511	1	0.061	–
Likelihood Ratio	4.136	1	0.042	–
Fisher's Exact Test	–	–	–	0.044
Linear-by-Linear Association	4.050	1	0.044	–
N of Valid Cases	300			

a. 0 cells (0.0%) have an expected count of less than 5. Minimum expected count = 43.50
b. Computed only for a 2x2 table

Table 10
Symmetric Measures

Measure	Value	Approx. Sig.
Phi	.117	.042
Cramer's V	.117	.042
N of Valid Cases	300	

Grammarly, by contrast, appeared more useful for localized and surface-level corrections such as punctuation-related clarity, concise expression, and some mechanical grammar suggestions. This distinction is important because it shows that the difference between the two tools is not simply a matter of interactivity, but of how each tool processes grammatical problems. ChatGPT's generative and explanatory feedback appears more beneficial for errors that require contextual reasoning, while Grammarly remains useful for quick proofreading and surface-level editing. The results, based on 300 grammar error cases, revealed a statistically significant association between the tool used and correction success, $\chi^2(1, N = 300) = 4.119$, $p = .042$. This indicates that the use of ChatGPT or Grammarly significantly influenced the likelihood of an accurate correction of an error.

As shown in the crosstabulation, ChatGPT successfully corrected 112 of 150 errors, while Grammarly corrected 94 of 150. These frequencies deviated from the expected counts under the null hypothesis, where both tools would have been equally effective (expected counts: 106.5 and 99.5 for ChatGPT and Grammarly, respectively). Additionally, the Likelihood Ratio Chi-Square and Linear-by-Linear Association also showed significant values ($p = .042$ and $p = .044$, respectively), reinforcing the strength of this association. Although the Continuity Correction was slightly above the threshold ($p = .061$), Fisher's Exact Test also confirmed statistical significance ($p = .044$).

The effect size, measured using Cramer's V and Phi coefficients (both = 0.117, $p = .042$), suggests a small but meaningful effect, implying that ChatGPT has a modest advantage over Grammarly in effectively handling a wider range of grammatical error types.

RQ3: How does the use of ChatGPT and Grammarly influence EFL students' academic writing development over time?

Table 11
Thematic Analysis Table – EFL Students' Writing Development Using ChatGPT and Grammarly

Theme	Sub-theme	Tool	Description	Representative Quote
Awareness of Grammar Rules	Explicit learning through explanations	ChatGPT	Students reported increased awareness of grammar rules through feedback	"ChatGPT explained to me why my sentence was wrong, so I remembered the rule."
	Passive correction	Grammarly	Grammarly often corrected errors silently, without a detailed explanation	"It corrected my grammar, but I didn't always understand why."
Writing Confidence	Increased self-revision skills	ChatGPT	Learners felt more confident revising drafts after interacting with ChatGPT.	"Now I feel more confident editing my essays after using ChatGPT."
	Dependence on the suggestions	Grammarly	Some students reported over-reliance on Grammarly's instant corrections	"I just clicked "accept," but I'm not sure if I really learned it."
Vocabulary and Style	Lexical variation	ChatGPT	ChatGPT offered richer vocabulary and sentence restructuring suggestions	"It helped me write more academically by suggesting better phrases."
	Concise expression	Grammarly	Grammarly improved clarity through word economy and punctuation corrections.	"It helped me make my writing shorter and more precise."
Motivation to improve	Interactive engagement	ChatGPT	The conversational style encouraged students to ask questions and revise their answers.	"It feels like a tutor. I can ask follow-up questions."
	Quick, surface-level feedback	Grammarly	Grammarly motivated revision, but less depth in learning	"It's useful for quick checking, but not for deep learning."
Autonomy and Learning Style	Encourages self-directed learning	ChatGPT	Promoted reflection and experimentation through writing	"I started thinking more about how to improve my

Theme	Sub-theme	Tool	Description	Representative Quote
				sentences."
	Tool dependency	Grammarly	Risk of surface-level corrections without internalization	"I used it for fast checking, but I didn't think much about the rules."

The reflection data further suggest that students' interaction with ChatGPT involved a more recursive learning process. Rather than simply accepting corrections, several students reported asking follow-up questions, requesting alternative sentence structures, comparing revised versions, and using explanations to revise similar errors in later drafts. This pattern indicates a shift from passive correction to active feedback engagement. In contrast, Grammarly was more often used as a confirmation and proofreading tool, where students accepted or rejected suggestions quickly without extended interaction. This difference helps clarify the learning process behind the quantitative results: ChatGPT supported explanation-based revision and metalinguistic noticing, while Grammarly supported efficiency-oriented editing. The ChatGPT was consistently associated with deeper cognitive engagement and metalinguistic awareness. For instance, students frequently mentioned how ChatGPT's explanations helped them understand and remember grammar rules, thereby supporting explicit learning. In contrast, Grammarly was more often linked to passive correction, where students accepted changes without fully understanding the rationale.

In terms of writing confidence, ChatGPT encouraged students to revise their drafts with greater independence, describing the experience as similar to working with a tutor. Conversely, some learners noted a tendency to become overly dependent on Grammarly's instant suggestions, potentially limiting their growth in self-editing skills. Regarding vocabulary and style, ChatGPT was praised for offering a richer academic tone and alternative phrasing, while Grammarly was valued for making writing more concise and clear.

The tools also differed in how they motivated students: ChatGPT's interactive, conversational nature engaged learners in a reflective process, whereas Grammarly was seen as a practical tool for quick corrections but less beneficial for long-term development. Lastly, ChatGPT appeared to foster autonomous learning, encouraging experimentation and conscious revision. Grammarly, while efficient, sometimes led to tool dependency, with students relying on automatic corrections without internalizing grammar rules.

Discussion

The primary aim of this study was to compare the effectiveness of ChatGPT and Grammarly in supporting the academic writing development of EFL university students, particularly in relation to corrective feedback, grammar accuracy, and perceived learning value. The findings revealed notable differences in the impact of the two AI tools impacted students' writing experiences through a combination of statistical analysis and thematic exploration. In response to RQ1, descriptive and inferential data indicated that students perceived ChatGPT as more effective than Grammarly in improving grammar, increasing revision confidence, and helping them understand grammar rules. Although both tools were rated similarly for feedback clarity, ChatGPT's ability to offer detailed explanations appeared to enhance students' sense of learning and engagement. Regarding RQ2, the Wilcoxon Signed-Rank Test confirmed that ChatGPT provided significantly more accurate grammar corrections than Grammarly across most error types. The Chi-square test further supported this, showing a statistically significant association between tool use and correction success. Finally, the thematic analysis for RQ3 highlighted five key dimensions of development—awareness of grammar rules, writing confidence, vocabulary and style, motivation, and autonomy—where ChatGPT was perceived to provide deeper support than Grammarly.

Collectively, these findings contribute to the understanding of how different AI tools function not only as writing assistants but also as pedagogical agents in the context of EFL instruction.

The findings of this study resonate with and expand upon a growing body of research highlighting the pedagogical value of AI-mediated feedback in second language writing. Previous studies on Grammarly (Ismailia & Novawan, 2025; Prasetya et al., 2024; Paff, 2023) have consistently shown that the tool is effective in detecting surface-level grammatical and mechanical errors, offering clear and immediate corrections related to punctuation, spelling, and sentence structure. This aligns with the results of the current study, where Grammarly was perceived as particularly useful for clarity and conciseness. However, Giray (2024) noted a lack of depth in learning due to Grammarly's tendency to provide corrective suggestions without meaningful elaboration—a limitation echoed in student reflections from this study, where some participants admitted to accepting Grammarly's suggestions without fully understanding the underlying rules.

In contrast, the present findings reveal that ChatGPT's interactive, dialogic feedback format significantly enhances learners' perceived understanding of grammar and writing skills. This supports the recent literature. Godwin-Jones (2024), Creely (2023); Muntazar and Hajar (2025) This highlights the educational potential of generative AI tools in language learning, primarily when they serve not just as correctors but also as facilitators of metalinguistic awareness. The ability of ChatGPT to engage in conversational exchanges allows learners to ask for clarification, request alternative suggestions, and explore explanations in real-time—features that are largely absent in fixed-output systems like Grammarly. Such affordances enable what Perifanou and Economides (2025) described as "language-related episodes"—moments in which learners pause to reflect on and negotiate language use—which are critical for internalizing grammatical rules.

Additionally, the noticing hypothesis VanPatten & Smith (2022), which posits that learners must consciously notice linguistic features in input to acquire them, provides a theoretical lens for interpreting these results. ChatGPT's elaborative feedback, including explanations and reformulations, arguably promotes this noticing process more effectively than Grammarly's minimalistic prompts. Learners' reported increases in confidence and awareness after using ChatGPT suggest that the tool supported conscious grammar learning, a core component of successful second language writing development. These findings are also consistent with research by Aladini et al. (2025) and Dong (2024), who found that ChatGPT encourages deeper engagement with language and fosters student reflection during writing tasks.

One possible explanation for ChatGPT's superior performance lies in its generative feedback design, which encourages student engagement beyond simple error correction. By providing contextual explanations, rephrasing options, and the opportunity for follow-up interaction, ChatGPT promotes deeper cognitive processing of grammar-related issues, which may enhance retention and transfer. In contrast, Paff (2023) explored that Grammarly's interface is optimized for efficiency, offering limited explanation and minimal learner interaction, which may account for its relatively lower perceived learning value. Another explanation is the role of user agency. Roffarello et al. (2025) explicated that ChatGPT's chat-based format allows learners to steer the direction of feedback, which may foster greater autonomy and motivation. In contrast, Grammarly provides fixed suggestions that are more likely to be passively accepted. However, Chen (2025) confirmed that Grammarly scored slightly higher in feedback clarity, likely due to its streamlined interface and one-click suggestions, which are particularly appealing for quick revision tasks. This highlights that tool preference may also be task-dependent: learners may benefit from using Grammarly for final-stage proofreading and ChatGPT during the drafting and revision stages that require deeper learning. This finding strengthens the originality of the present study because it specifies the nature of ChatGPT's advantage rather than merely repeating the common claim that ChatGPT is more interactive than Grammarly. The data suggest that ChatGPT's

pedagogical value is most evident when learners deal with grammar problems that require explanation, contextual interpretation, and reformulation. In contrast, Grammarly's value appears more prominent in rapid proofreading and surface-level refinement. Therefore, the contribution of this study lies in mapping tool effectiveness according to feedback function and grammar category, rather than treating AI writing tools as uniformly effective or ineffective. The small yet significant effect size observed in the grammar correction data (Cramer's V = 0.117) suggests that, although the advantage of ChatGPT is not dramatic, it is consistent and meaningful, especially when considered alongside students' reflections and reported learning experiences.

The observed superiority of ChatGPT in both perceived learning value and correction accuracy can be attributed to several interconnected factors, most notably its generative feedback design, which more closely aligns with the principles of second language acquisition (SLA) and writing pedagogy. Unlike Grammarly, which provides fixed, often surface-level suggestions, ChatGPT engages learners in dynamic interactions where feedback is not only offered but also questioned, clarified, and expanded upon. This interactivity may activate higher-order cognitive processes, such as analysis, evaluation, and synthesis—core components of deeper learning. For example, when students ask ChatGPT *why* a sentence is incorrect or request alternative phrasings, they engage in metalinguistic reflection, a key driver of long-term language development. This aligns with Matusov et al. (2024) Sociocultural theory highlights the role of dialogue in constructing knowledge. In this context, ChatGPT acts as a form of scaffolded peer or tutor within a learner's Zone of Proximal Development (ZPD).

These findings also highlight the ethical dimension of AI-assisted writing development. The reflection data suggest that ChatGPT encouraged more active engagement because students could ask follow-up questions, request explanations, and compare alternative formulations. This supports the view that generative AI may enhance learner autonomy when it is used as a dialogic feedback resource rather than as a text-production shortcut. However, the study also shows that AI tools may create dependence when students accept suggestions without evaluating the linguistic rationale behind them. This issue was particularly visible in students' responses to Grammarly, where some learners reported accepting corrections quickly without fully understanding the underlying grammar rules. Therefore, the pedagogical integration of AI writing tools should include explicit instruction on academic integrity, responsible AI use, and feedback literacy. Students should be trained to treat AI-generated feedback as input for reflection and revision, not as an automatic substitute for their own writing decisions.

CONCLUSIONS

The study contributes to AI-assisted EFL writing research by demonstrating that the difference between ChatGPT and Grammarly is not limited to general interactivity or user preference, but extends to grammar-category performance, feedback function, and learner engagement patterns. ChatGPT showed stronger pedagogical value in feedback situations requiring explanation, contextual reformulation, and reflective revision, whereas Grammarly remained useful for quick proofreading, clarity, and surface-level refinement. The findings point to meaningful distinctions in how each tool contributes to grammar awareness, learner autonomy, and writing confidence. While both tools served as helpful writing aids, their pedagogical value appeared to diverge in terms of feedback depth, interactivity, and learner engagement. The results carry several implications for language educators and curriculum developers. Instructors may consider integrating tools like ChatGPT not merely as revision aids but as interactive feedback resources that encourage reflection and self-correction during the writing process. Meanwhile, Grammarly may continue to serve a role in final-stage editing, particularly for checking mechanical accuracy and improving clarity.

The findings suggest that ChatGPT can be used as an interactive feedback resource during the drafting and revision stages of EFL academic writing. Unlike tools that only provide direct corrections, ChatGPT allows students to ask follow-up questions, request explanations, and compare alternative sentence structures. This interaction may encourage learners to reflect on their grammatical errors and engage in self-correction during the writing process. Therefore, teachers may guide students to use ChatGPT not as a tool for generating complete texts, but as a learning aid for understanding feedback, revising sentences, and improving grammar awareness. Meanwhile, Grammarly may be more appropriate for the final proofreading stage, particularly for checking correctness, clarity, and conciseness. In this way, both tools can be integrated in a phase-specific manner, with ChatGPT supporting reflective revision and Grammarly supporting surface-level polishing.

Despite its contributions, this study is not without limitations. The sample was drawn from a specific university context, which may limit the generalizability of the findings to other educational or cultural settings. Additionally, the data collection focused on short-term tool usage during a limited intervention period. A longer-term study could provide insights into the sustained impact of AI-assisted writing tools on language development and learner autonomy. Future research could extend this work by exploring how these tools influence revision strategies over time or by examining the interaction between learner proficiency level and tool effectiveness. Investigating how AI-generated feedback interacts with teacher feedback and whether blended approaches offer complementary benefits would be another fruitful direction.

REFERENCES

Aladini, A., Ismail, S. M., Khasawneh, M. A. S., & Shakibaei, G. (2025). Self-directed writing development across computer/AI-based tasks: Unraveling the traces on L2 writing outcomes, growth mindfulness, and grammatical knowledge. *Computers in Human Behavior Reports*, *17*, 100566. https://doi.org/10.1016/j.chbr.2024.100566

Algburi, E. A., & Razali, A. B. (2022). Role of feedback on English academic writing skills of tertiary level Iraqi English as a foreign language (EFL) students: A review of literature. *Journal of Academic Research in Progressive Education and Development*, *11*(1), 689–702. https://doi.org/10.6007/IJARPED/v11-i1/12168

Alostath, K. (2021). Graduate students' challenges in academic writing. A Thesis Proposal.

Amina, F. (2024). *Exploring EFL students 'perceptions on the use of Grammarly as an AI writing tool to enhance academic writing proficiency: The case study of Master Two English students at Mohammed Kheider University* [Master's thesis, Mohammed Kheider University].

Asadi, M., Ebadi, S., & Mohammadi, L. (2025). The impact of integrating ChatGPT with teachers' feedback on EFL writing skills. *Thinking Skills and Creativity*, *56*, 101766. https://doi.org/10.1016/j.tsc.2025.101766

Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, *11*(4), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

Chen, Q. (2025). Students' perceptions of AI-powered feedback in English writing: Benefits and challenges in higher education. *International Journal of Changes in Education*. https://doi.org/10.47852/bonviewIJCE52025580

Creely, E. (2023). The possibilities, limitations, and dangers of generative AI in language learning and literacy practices. *International Graduate Research Symposium 2023. Proceedings*, 1–14.

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. https://doi.org/10.1007/s12144-024-05728-9

Giray, L. (2024). "Don't let Grammarly overwrite your style and voice:" Writers' advice on using Grammarly in writing. *Internet Reference Services Quarterly*, *28*(3), 293–303.

https://doi.org/10.1080/10875301.2024.2344762
Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *ArXiv Preprint ArXiv:2403.20216*. https://doi.org/10.48550/arXiv.2403.20216
Iorliam, A., & Ingio, J. A. (2024). A comparative analysis of generative artificial intelligence tools for natural language processing. *Journal of Computing Theories and Applications*, *1*(3), 311–325. https://doi.org/10.62411/jcta.9447
Ismailia, T., & Novawan, A. (2025). Enhancing EFL students' recount writing: A comparative analysis of Grammarly, QuillBot, and GrammarCheck applications. *Journal of English in Academic and Professional Communication*, *11*(2), 89–102.
Kot, S. O., & Nykyporets, S. S. (2025). Activating students cognitive engagement in technical English learning with AI tools. *Science and Education in the Third Millennium: Information Technology, Education, Law, Psychology, Social Security and Work, Management: International Collective Monograph. Vol. I. Lublin, Polska. Chap. 9: 295-332.* https://doi.org/10.5281/zenodo.16942267
Ladeinde, O. J. (2023). *An adaptive feedback framework for a language-independent intelligent programming tutor (IPT) using ANTLR*.
Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*(1), 50. https://doi.org/10.1186/s40561-024-00304-0
Mali, Y. C. G. (2025). Exploring the use of ChatGPT in EFL/ESL writing classrooms: A Systematic Review. *Journal of Language and Education*, *11*(2 (42)), 137–156.
Matusov, E., Gpt, C., Smith, M. P., & Shugurova, O. (2024). Does ChatGPT4 have a dialogical self?: A Bakhtinian perspective. *Culture & Psychology*, *30*(4), 841–870.
Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: a study on proficiency, revision frequency and writing quality. *Discover Education*, *4*(1), 170.
Mun, C. (2024). EFL learners' English writing feedback and their perception of using ChatGPT. *Journal of English Teaching through Movies and Media*, *25*(2), 26–39. https://doi.org/10.16875/stem.2024.25.2.26
Muntazar, T., & Hajar, I. (2025). Discursive and metalinguistic perspectives on AI-generated writing feedback: A student perception study. *Journal Informatic, Education and Management (JIEM)*, *7*(2), 516–526.
Paff, J. B. (2023). *Cognitive feedback theories and artificial intelligence: A case for a Grammarly of UI/UX design*.
Perifanou, M., & Economides, A. A. (2025). Students collaboratively prompting ChatGPT. *Computers*, *14*(5), 156. https://doi.org/10.3390/computers14050156
Prasetya, R. E., Hariyadi Didik Hariyadi, R., & Sofiani, F. (2024). Identifying Grammarly automated writing evaluation and self-editing by administrative officer. *JELLT (Journal of English Language and Language Teaching)*, *8*(1), 120–133.
Roffarello, A. M., Calò, T., Scibetta, L., & De Russis, L. (2025). Investigating how computer science researchers design their co-writing experiences with AI. https://doi.org/10.1145/3706598.3713205
Saito, K. (2021). *Effects of corrective feedback on second language pronunciation development*. Cambridge University Press.
Shadiev, R., & Yang, M. (2020). Review of studies on technology-enhanced language learning and teaching. *Sustainability*, *12*(2), 524. https://doi.org/10.3390/su12020524
Taj, S., & Khan, M. A. (2024). Comparing Grammarly and ChatGPT for automated writing evaluation of ESL learners. *Bahria University Journal of Humanities & Social Sciences*, *7*(2), 63–90.
Takallou, F. (2025). The impact of Grammarly AI-powered writing tool on the EFL Students' Writing Proficiency and Autonomy. *Technology Assisted Language Education*, *3*(2), 99–119.

Tan, X., Reynolds, B. L., & Ha, X. Van. (2024). Oral corrective feedback on lexical errors: a systematic review. *Applied Linguistics Review*, *15*(3), 1177–1221.

VanPatten, B., & Smith, M. (2022). *Explicit and implicit learning in second language acquisition*. Cambridge University Press.

Wang, C., Li, Z., & Bonk, C. (2024). Understanding self-directed learning in AI-Assisted writing: A mixed methods study of postsecondary learners. *Computers and Education: Artificial Intelligence*, *6*, 100247. https://doi.org/10.1016/j.caeai.2024.100247

Zotzmann, K., & Sheldrake, R. (2021). Postgraduate students' beliefs about and confidence for academic writing in the field of applied linguistics. *Journal of Second Language Writing*, *52*, 100810. https://doi.org/10.1016/j.jslw.2021.100810