BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Effectiveness of Accurate ChatGPT Commands in Enhancing A2 Level Academic Writing Feedback

Jhomayra Isabel Hoyos Valdiviezo [1*] **, *Jardel Coutinho dos Santos*** [1] **, *Rafaela María Vinueza Beltrán*** [2]

[1] *Educational Science Department, Universidad Técnica Estatal de Quevedo, Quevedo 120301, Ecuador*

[2] *Educational Science Department, Universidad Yachay Tech, Urcuqui 100115, Ecuador*

ABSTRACT

Academic writing presents significant challenges for beginner learners of English as a Foreign Language (EFL), where learners tend to make mistakes in grammar, coherence, and structure. Feedback is essential for the development of writing, yet in traditional instructional models it is not always provided with the quality or frequency required. This study explores the effectiveness of ChatGPT-generated feedback in enhancing academic writing skills among A2-level English as a Foreign Language learners. Involving a pre-post experimental design, 45 students participated over a fourteen-week intervention period within two groups: one receiving traditional instruction and the other using ChatGPT guided by a tailored prompt created by the principal researcher: *"Evaluate this paragraph based on Cambridge A2-level standards (CEFR) for Content, Organization, and Language (vocabulary and grammar). Provide specific feedback"*. The results showed statistically significant gains in writing performance for the experimental group, especially in language accuracy and structural organization, with no regression in any skill area, confirming that AI feedback has the potential to facilitate language acquisition among participants at a beginner level, without compromising learner autonomy, when conducted through a structured approach. The research indicates that ChatGPT can be utilized as a pedagogical tool in addition to regular education, serving as a scalable resource with the aid of clear evaluative criteria. Future research is encouraged to expand this approach to other CEFR proficiency levels and explore learners' qualitative experiences with AI feedback.

*CORRESPONDING AUTHOR:

Jhomayra Isabel Hoyos Valdiviezo, Educational Science Department, Universidad Técnica Estatal de Quevedo, Quevedo 120301, Ecuador; Email: jhoyosv2@uteq.edu.ec

ARTICLE INFO

Received: 14 August 2025 | Revised: 25 August 2025 | Accepted: 11 September 2025 | Published Online: 3 November 2025
DOI: https://doi.org/10.30564/fls.v7i12.11647

CITATION

Hoyos Valdiviezo, J.I., Santos, J.C.d., Vinueza Beltrán, R.M., 2025. Effectiveness of Accurate ChatGPT Commands in Enhancing A2 Level Academic Writing Feedback. Forum for Linguistic Studies. 7(12): 12–25. DOI: https://doi.org/10.30564/fls.v7i12.11647

Highlights

- The first quasi-experimental study used structured ChatGPT prompts to improve A2-level academic writing in EFL settings.
- Inter-rater reliability validates the statistically significant increases in grammar, vocabulary, and organization demonstrated.
- Provide a feasible 12-week intervention framework that teachers can incorporate into their weekly lesson plans.
- Emphasize the balance between AI feedback and human mediation, supporting the pedagogical foundation.
- Provides implications for extending AI-assisted feedback across multiple CEFR competency levels and investigating cognitive processes involved in L2 writing.

Keywords: AI; Academic Writing; ChatGPT; CEFR; EFL; Feedback; Prompt; Writing

1. Introduction

Effective academic writing is a fundamental skill for students of English as a foreign language (EFL) at all levels; however, learners at an A2-level students often face persistent challenges with grammar, coherence, and clarity [1]. Attributing these difficulties solely to limited linguistic knowledge such as grammar and vocabulary would be misleading. Writing proficiency depends on a range of communicative abilities, and although insufficient exposure to the conventions of academic writing certainly restricts students' capacity to produce coherent texts, it is not the only factor at play.

Due to this, the role of the teacher becomes essential, as they traditionally provide feedback that is valuable for establishing an effective writing process. However, due to multiple factors, this feedback can be delayed, inconsistent, and irregular, which in turn can discourage learning, as students do not have the opportunity to make corrections and prompt adjustments [2]. Given these limitations, another pedagogical method or requirement for offering alternative feedback mechanisms is to help provide timely, structured, and personalized support.

Recent advancements in artificial intelligence (AI) technology have introduced AI-powered writing assistants, such as ChatGPT, which present potential opportunities to address these pedagogical limitations. ChatGPT's ability to generate instant, contextually responsive feedback allows learners to identify and correct errors in grammar, structure, and consistency more rapidly than is typically possible through conventional teacher feedback [3]. Prior studies suggest that targeted, real-time suggestions provided by AI systems may facilitate iterative refinement of writing skills. However, while such findings are promising, they should be interpreted cautiously, as AI-based feedback may not fully replace teacher guidance. Specifically, AI tools are not able to assess more extensive rhetorical or argumentation apparatus, and their level of feedback may differ depending on the quality of the input [4].

Besides, it is noted that the systematic and well-written prompt design has become an essential tool in the optimization of ChatGPT as an instructional tool. On the other hand, depending heavily on AI can be one of the risks that students may develop excessive dependency on the educational assistive tool, thus curtailing active participation in the writing process [2]. A pedagogical approach that integrates AI feedback within a framework encouraging critical evaluation and learner autonomy is therefore essential [3]. Although research into AI-assisted feedback has expanded in recent years, there is limited empirical evidence on its application for teaching EFL writing specifically to learners at an A2 level. Moreover, the effect of structured ChatGPT prompts—aligned with established writing standards such as those of Cambridge—in the development of grammatical accuracy, lexical precision, and textual coherence remains underexplored. Addressing this gap is vital to determine whether AI-generated feedback can complement traditional instruction and measurably improve students' writing proficiency.

To contribute to this area of research, this study examines the effectiveness of feedback provided by ChatGPT as a method for improving the academic writing of A2-level EFL students. By analyzing students' writing, ChatGPT measures skills based on a well-structured prompt that adheres to Cambridge international standards and generates effective

feedback that provides students with instant, consistent, and regular analysis of their strengths and weaknesses.

Specifically, the research examines whether precisely formulated prompts improve grammatical accuracy, vocabulary use, coherence, and overall writing proficiency. To achieve this, a quantitative research design was used to measure the impact of AI-generated feedback on student performance, tracking progress before and after exposure. Accordingly, the research addressed the following question: How effective are well-designed ChatGPT prompts in providing feedback that meets Cambridge standards to enhance the writing skills of A2 learners?

The findings concluded that well-designed ChatGPT prompts, when integrated into a well-organized pedagogical system, can lead to measurable growth in linguistic accuracy and structural integrity in the writing process among A2-level EFL learners.

1.1. Literature Review

1.1.1. Challenges in EFL Academic Writing

Linguistic, cognitive, and pedagogical barriers make the task of academic writing a significant challenge to the learners of English as a Foreign Language (EFL). Students struggle with their grammar, coherence, and cohesion, making it challenging to produce well-structured texts [1,2]. These issues stem from native language interference, limited exposure to academic English, and inadequate instruction [2,4]. Many studies indicate that syntactic and morphological accuracy is a key difficulty; for instance, Pasaribu et al. [1] found that Indonesian EFL students frequently make syntactic errors (61.69%), followed by punctuation and capitalization mistakes (23.37%). Moreover, interlingual interference plays a significant role when students use rules of their native language in English, and thus, writing loses clarity [1,5] and the vocabulary is repetitive [2,6], which also affects the overall quality of writing. In addition, recent research has indicated that not only does this linguistic constraint hinder the correctness of grammar, but it also inhibits the capacity to deal with a sophisticated sentence layout and sustain a coherent argument throughout the paragraphs, especially in the education setting where the students have not much exposure to English [7].

On the same note, there are also difficulties in cohesion and coherence. Brilliant writing presupposes a logical structure and transition, but many learners do not observe a logical chain of ideas [2,5]. References, citations, and argument structure are some of the challenges that students face in writing in higher education in Pakistan [2,8]. Krause et al. [8] further note that these organizational problems are often exacerbated by the tendency to translate directly from the L1, which slows down writing speed, reduces lexical variety, and creates disjointed discourse structures.

Pedagogical aspects, on the other hand, also contribute to these challenges. The response quality and proficiency of teachers have a considerable influence on student development [2,4]. Most of them do not receive any training in composition education, instead learning about grammar and spelling rather than argument and analysis [2,7]. Writing is also determined by motivation and self-efficacy. Negative comments, low self-confidence, and fear of errors contribute to writing anxiety, which can lead to procrastination or failure to complete assignments among many learners [4,5].

Similarly, second language writing is a complex cognitive process that involves planning, monitoring, and revising, and this process alone consumes a significant portion of working memory space. This is because EFL learners have a low retention ability, which hinders their ability to sustain syntactic complexity and lexical heterogeneity, as well as general text coherence, due to their limited working memory [9]. The restraints highlight the factors that make it challenging for beginner writers to control both grammar and word structure simultaneously. Recent evidence suggests that one can utilize ChatGPT to enable superficial correction; however, the issue of cognitive overload in learners also still exists, and intervention targeted at minimizing the working memory burden during writing should also have a higher priority as an enabler [9].

Research indicates that such anxiety is compounded by the slow, irregular nature of traditional teacher feedback, which often reduces students' willingness to revise their work and engage in multiple drafting cycles [8,9]. The explosive mindset and positive feedback may reduce such fears [6,8]. It is necessary to address these obstacles using clear instructions, direct feedback, and engaging learning activities [2,4]. AI-powered tools, such as ChatGPT, have shown potential to address part of this gap by offering immediate, tailored feedback that helps learners identify and correct errors in

grammar, vocabulary, and organization [7,8]. However, multiple authors stress that these tools should be implemented within blended pedagogical models, where teachers provide targeted instruction on higher-order skills—such as argumentation, rhetorical coherence, and discipline-specific writing conventions—that AI cannot yet fully evaluate [7].

The issues that require further research include blended practices in which teachers provide instruction and AI offers assistance. In particular, more empirical work is needed to determine how such integrated approaches can be optimized to improve both surface-level accuracy and deep writing skills, while also mitigating risks of overdependence on AI tools [8].

1.1.2. Understanding AI Literacy

With the increasing role of AI technologies in educational practice and professional environments, AI literacy has become one of the key competencies of teachers and students. It involves not only the knowledge of AI systems and how they work, but also an analysis of their use, their shortcomings, and moral applications [10]. AI literacy requires not just basic functional knowledge but also consideration of the trustworthiness and reliability of AI-generated content, as uncritical consumption creates misinformation, overreliance, and perpetuation of pre-existing biases [2,11].

A prominent feature of AI literacy is prompt engineering, i.e., formulating queries to maximize the relevance, precision, and teaching value of AI answers [3]. Research indicates that structured and guided prompting substantially improves the accuracy and pedagogical usefulness of AI outputs, reducing errors and irrelevant suggestions [11,12]. This is especially significant where EFL is concerned, as immediate clarity may directly impact the learning of the language by making the learning AI more inclined towards level-suitable corrections and examples.

Moreover, AI literacy must address the ethical dimensions of AI-assisted learning, including risks related to data privacy, algorithmic bias, and academic integrity [4,12]. For example, Kusuma et al. [12] found that preservice teachers with formal AI literacy training were better able to customize AI-generated lesson plans for diverse learner needs while mitigating ethical risks. Similarly, Ng et al. [10] demonstrated that AI tools can enhance self-regulated learning through adaptive feedback, provided that students are trained to critically evaluate and refine AI suggestions rather than accept them passively.

In addition, studies in higher education reveal that AI literacy training can increase student engagement and writing motivation by helping learners see AI as a collaborative partner rather than a replacement for cognitive effort [9,12]. Recent findings also emphasize that AI literacy fosters metacognitive skills, enabling learners to plan better, monitor, and evaluate their work [8]. However, without explicit instruction on the limitations of AI models—such as their inability to fully grasp rhetorical nuance or cultural context—students may misinterpret feedback, leading to surface-level revisions without deeper improvement in writing quality [2,3].

Also, some findings in EFL contexts indicate that students' writing difficulties, such as weak grammar competence, lack of coherence, and limited vocabulary, could be partially mitigated by AI when used within a structured pedagogical framework that integrates human feedback [4]. Nonetheless, excessive use of AI without any critical literacy can help to maintain generic writing tendencies, as Mahmood [2] states that technological tools are not meant to replace but facilitate the acquisition of higher levels of writing development.

Accordingly, AI literacy is more of a coherent development that incorporates technical learnings of AI tools, the tactical design of prompts, engagement with ethical considerations, and a reflective use strategy as pedagogy. Since it is a comprehensive preparation, both educators and students can use the power of AI, retain academic integrity, promote autonomy, and protect ethical principles.

1.1.3. ChatGPT-Assisted Feedback

Artificial intelligence (AI) as a tool for academic writing feedback has gained momentum in recent years due to its capacity to provide automated, prompt, and personalized responses that complement or partially substitute traditional approaches [10]. In the English as a Foreign Language (EFL) context, ChatGPT has been recognized for delivering formative assessment feedback that targets grammar, structure, and lexical accuracy in real time, addressing one of the main limitations of human feedback—its delayed nature due to time and workload constraints [13]. Instant feedback promotes an iterative revision, fosters self-guided learning, and encourages students to take ownership of the editing process [8,14].

According to empirical evidence, ChatGPT shows strong performance when evaluating organization, coher-

ence, and grammatical accuracy, with other works pointing out that the recent revisions of the model are superior to the previous models in stability and accuracy [15]. It is scalable and accessible, meaning that it can be used as a resource in different educational contexts where human feedback is often unavailable or irregular [16]. In addition to the correction of errors, ChatGPT may facilitate self-regulated learning by allowing students to identify patterns in their mistakes and to develop revision techniques [10].

Nevertheless, notable limitations persist. Higher-order writing skills, including goals of development of an argument, rhetorical appeal, and critical interaction with ideas, are frequently not served by AI feedback [8,17]. Furthermore, it cannot replicate the communicative and affective dimensions of teacher feedback, which are essential for fostering motivation and sustained learner engagement [9]. Misinterpretation of AI suggestions is another documented challenge, as some students struggle to apply corrections accurately without guidance [14]. To avoid these problems, the importance of feedback literacy should be stressed, as it is a skill set that would help the learner to critically assess the input generated by AI before incorporating it into their work [10].

From a cognitive perspective, AI-assisted feedback can also serve as a mitigation of the cognitive burden associated with sensing and correcting errors. By shifting these lower-level operations outside, ChatGPT feedback allows learners to redirect working memory resources to more upper-level aspects of composing, such as argument construction and text organization. This is corroborated by studies from the higher education context where students using ChatGPT claim that the automation of routine tasks frees up more effort to engage with more demanding, critical tasks related to their academic writing and thinking activity [8].

Pedagogically, the idea of combining ChatGPT into an educational scheme better suits a hybrid model whereby the elements of artificial intelligence supplement instead of substitute for human mediation. The role of a teacher cannot be underrated in introducing AI feedback into context, facilitating the organization of complex revisions, and ensuring that learners acquire not only linguistic correctness but also higher-order cognitive and rhetorical skills [14,16]. Structured prompt engineering is also essential; high-quality, precise prompts have been shown to enhance the relevance and accuracy of ChatGPT's feedback significantly [10].

To conclude, although the ChatGPT machine helps users derive scalable, consistent, and immediate feedback that can be used to enhance EFL writing teaching and learning, it is only valid when administered alongside a practical pedagogical framework. Such integration needs to focus on teacher mediation, feedback literacy, and strategic prompting to capitalize on its advantages and address its shortcomings in terms of creating a sustainable and balanced approach towards EFL writing development [8,10].

2. Materials and Methods

2.1. Research Design

This research employed a quantitative study to assess the effectiveness of ChatGPT feedback in enhancing the basic academic writing skills of A2 learners of the English language. A quasi-experimental approach that included a pre-test and a post-test was implemented, as it allows for the measurement of individual student progress over time, offering an objective assessment of learning gains [18,19]. Unlike the more traditional outcome measures that only compare final course scores or that compare different cohorts, pre-post design follows the same people over time, so fewer confounding factors and stronger conclusions about the effectiveness of the instruction can be drawn [18].

In the context of writing instruction research, this design is particularly valuable because it allows the detection of subtle yet meaningful changes in learners' performance that might remain hidden when only comparing independent groups. Tesch [19] cautions that without longitudinal measurement of the same individuals, cohort differences may obscure actual learning gains, leading to inaccurate or even misleading conclusions. The specified methodological decision is further justified by the fact that it complies with the best practices of applied linguistics research since the within-subject change is more valuable pedagogically than using a cross-sectional assessment.

The experiment for this research was conducted over fourteen weeks, starting after the first midterm of the semester, with weekly 60-minute sessions. All participants completed pre- and post-tests assessed with the Cambridge rubric *"Writing Assessment subscales for A2 Key for Schools"* (Appendix A), which evaluates content, organization, and language on a 0–5 scale. During the twelve-week intervention

period, the experimental group completed short academic writing tasks and received immediate, tailored feedback via a structured ChatGPT prompt (Appendix B). The structured nature of the given prompt was aimed not only at making the feedback relevant and deep but also as a homework in AI literacy, making students learn how to engage with generative AI effectively and critically. The process facilitated feedback literacy because it helped learners be better equipped to understand, think, and apply the feedback produced by the AI instead of passively accepting it.

In contrast, the control group, followed the same syllabus and task sequence, but received traditional input consistent with standard classroom practices. By holding the instructional content constant and varying only the feedback delivery mechanism, the study could attribute the observed differences more confidently to the integration of AI-assisted feedback rather than to variations in instructional input. This design not only strengthens internal validity but also enhances the potential to generalize findings to other EFL classrooms considering technology-enhanced formative assessment.

2.2. Participants

The study involved 45 third-semester students from parallel A and B of the Language Pedagogy program of a public university in Ecuador. Parallel A (experimental group) consisted of 30 students (73.33% female, 26.67% male), and Parallel B (control group) was comprised of 15 students (66.67% female, 33.33% male), with ages ranging from 18 to 23 years.

2.3. Innovation

Traditional feedback strategies in EFL academic writing, especially for A2 learners, are often generic and delayed. Evaluations measuring student development through pre-post comparisons provide more valid outcomes than single assessments [19]. Thus, this study implemented innovative, real-time AI-powered feedback aligned with structured learning phases. In line with Jacobsen and Weber [20], the students were educated on using effective prompts with ChatGPT to obtain feedback on their academic writing. Such a practice aligns with the AI literacy pedagogy, where it is essential to recognize immediate engineering as a key determinant of digital literacy [16].

Participation in the training allowed the following guided activities: (1) use the prompt designed by the principal investigator: *"Evaluate this paragraph based on Cambridge A2-level standards (CEFR) for Content, Organization, and Language (vocabulary and grammar). Provide specific feedback."* (Appendix B), (2) receive feedback generated by ChatGPT, (3) revise writings having received the feedback based on the remarks made by ChatGPT, and (4) reflect on whether the feedback was helpful. As an extra step, each participant analyzed the feedback and transcribed it into their notes.

2.4. Intervention Description

The intervention lasted twelve weeks, with one 60-minute session per week. Both groups received instruction based on the same syllabus. Students wrote paragraphs on academic writing topics assigned weekly (descriptive, narrative, argumentative, expository), covering all stages of the writing process.

Experimental group participants photographed and submitted their paragraphs to ChatGPT using the predefined prompt (Appendix B). Before the first submission, students received a 45-minute training session on prompt usage, including step-by-step guidance on how to paste their text, interpret the generated feedback, and ask follow-up clarification questions if needed. Sample interactions were demonstrated, such as inputting a paragraph on climate change and receiving targeted suggestions for improving transitions and vocabulary.

Each week, the assigned topic was related to real-life academic contexts, for example, Week 1: Describe the image presented in class; Week 3: Narrate a memorable personal event; Week 6: Present an argument about pollution; Week 9: Explain the importance of learning English, etc. The required length of each paragraph in each intervention ranged from 100 to 150 words. The contributions were assessed for coherence and cohesion, grammatical accuracy, lexical variety, and adherence to the assigned text type, and the comments guided improvements in the following weeks. Each participant analyzed and transcribed the feedback generated into their notes, identifying both strengths and weaknesses. The researchers verified the correct use of the instructions after each activity and observed that those notes were used in subsequent writing tasks.

Participants of the control group completed the same tasks each week and received feedback on their work from the course instructor, in accordance with standard classroom practices. In addition, the instructor posted brief comments on the task reviews in the university's academic management system. These comments related to grammar, but were not analyzed under any rubric; instead, they were based on the requirements of the assigned task.

Pre-tests and post-tests consisted of a 100 to 150-word academic paragraph written under timed conditions (60 minutes) in a controlled environment. Both groups completed these assessments on paper under the supervision of the principal researcher. **Figure 1** provides a schematic overview of the study design.

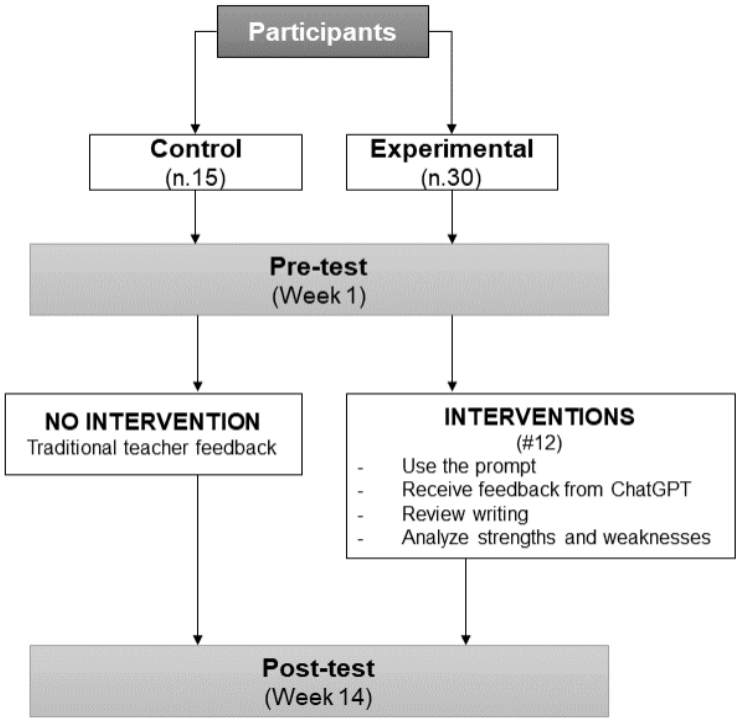

Figure 1. Overview of the experimental design for AI-assisted writing feedback.

2.5. Instruments

Writing performance on pre-tests and post-tests was assessed using an adapted Cambridge A2 writing rubric, evaluating Content, Organization, Vocabulary, and Grammar[21]. The rubric (Appendix A) included a scoring range from 0 to 5 for each criterion.

2.6. Data Collection and Analysis

All writing samples were scored independently by two trained raters using the Cambridge "Writing Assessment Subscales for A2 Key for Schools" rubric (Appendix A). Inter-rater reliability was assessed using Cohen's Kappa coefficient for each rubric category. Agreement levels were interpreted following Landis and Koch's[22] benchmarks, where values above 0.80 indicate almost perfect agreement, 0.61–0.80 substantial, and 0.41–0.60 moderate. The resulting Kappa values demonstrated substantial to almost perfect agreement across all criteria: content ($\kappa = 0.82$), organization ($\kappa = 0.78$), and language ($\kappa = 0.84$), confirming the consistency of scoring between raters.

Data were analyzed using IBM SPSS Statistics 27.0. Descriptive statistics (mean, standard deviation) were calculated for each group, followed by paired-sample t-tests to measure within-group progress, and independent-sample t-tests to compare between-group differences. Statistical significance was set at $p < 0.05$.

Before running parametric tests, the assumptions of normality and homogeneity of variances were verified. The Shapiro–Wilk test confirmed that score distributions did not significantly deviate from normality in either group, and Levene's test indicated homogeneity of variances across groups.

These results validated the use of paired and independent-sample t-tests for the present analysis. Otherwise, non-parametric alternatives (e.g., Wilcoxon signed-rank test or Mann–Whitney U test) would have been applied [23].

The independent variable was the use of an accurately designed ChatGPT prompt. In contrast, the dependent variable was an improvement in A2-level academic writing skills, measured through three rubric dimensions: content, organization, and language. Each of its standards was scored in bands from 0 to 5, where 0 and 1 indicate irrelevance to the task, and 5 represents the highest level of learning. The researchers assigned a band to each rubric standard and then averaged it across the three standards at the end of the scoring, based on each student's performance on the pre- and post-tests.

2.7. Ethical Considerations

The study strictly adhered to ethical guidelines, ensuring voluntary participation with informed consent and maintaining participant anonymity and data confidentiality. Ethical approval was granted by the Career Coordination Office of the university to which the participants belong. Participants were fully informed about the research and retained the right to withdraw at any time without consequences, according to the ethical guidelines for educational research.

2.8. Data and Materials Availability

The dataset generated and analyzed during this study is available in the Figshare repository at https://doi.org/10.6084/m9.figshare.29755598.v1 [24]. The data is in an anonymous format and ready for replicable use. The research instruments, such as the ChatGPT prompt and the rubric, are available in the appendices of this document.

3. Results

The author ran a paired t-test in SPSS to compare the pre- and post-test results of the experimental and control groups. **Table 1** demonstrates statistically significant improvement following the ChatGPT intervention. Specifically, scores increased from the pre-test mean (M = 2.94, SD = 0.50) to the post-test mean (M = 3.62, SD = 0.57). This improvement was confirmed by a paired-sample t-test, indicating a significant difference: t (34) = −8.03 and $p < 0.001$, with a large effect size (d = 0.46). The negative sign of the t-value indicates the direction of the output improvement between pre-test and post-test, and does not mean a negative influence.

On the other hand, the control group showed a non-significant difference between pre-implementation (M = 2.75, SD = 0.89) and post-implementation (M = 2.92, SD = 0.54) scores when examining all the samples together; t (17) = −1.35, $p = 0.19$, with a medium effect size (d = 0.57).

To provide credibility in the scoring procedure, inter-rater agreement was computed, yielding a value of 0.87, based on the Cohen Kappa coefficient. This outcome is significantly higher than the standard value of 0.80 [22], which corresponds to a high extent of consistency between the raters when rating the writing samples.

Additionally, **Table 2** demonstrates that the enhancements in the experimental group were interchangeable, as the average post-test mean was greater than the average pre-test mean in all rubric standards, including content, organization, and language. Notably, the negative changes in scores were absent, which implied that the intervention did not exert any negative changes. This means that the intervention did not have any adverse effects. The results would validate the hypothesis and highlight the ability of AI-based feedback systems (ChatGPT) to improve the quality of the learning process of programs in EFL contexts.

Table 1. Paired Sample T-test Results from Experimental Group and Control Group.

Experimental (N = 30)						
Test	**M**	**SD**	**Min**	**Max**	**Sig. (P)**	**Cohen's d**
Pre	2.94	0.50	1.66	4	<0.001	0.46
Post	3.62	0.57	2.66	5		
Control (N = 15)						
Test	**M**	**SD**	**Min**	**Max**	**Sig. (P)**	**Cohen's d**
Pre	2.75	0.89	1.33	4	0.19	0.57
Post	2.92	0.54	1.66	4		

Note: N = sample; M = mean; Sd = standard deviation; Sig (P): significance; Cohen's d = effect size.

Table 2. Descriptive Statistics of Rubric's Standards.

Experimental (N = 30)				
Standards	**Pre-Test Average**	**Post-Test Average**	**M**	**SD**
Content	3.86	4.14	4.00	0.19
Organization	2.93	3.50	3.21	0.40
Language	2.03	3.23	2.63	0.85
Control (N = 15)				
Standards	**Pre-Test Average**	**Post-Test Average**	**M**	**SD**
Content	3.40	3.66	3.53	0.18
Organization	2.66	3.00	2.83	0.24
Language	2.20	2.20	2.20	0.00

Note: N = sample; M = mean; Sd = standard deviation.

4. Discussion

The statistically significant improvement in writing scores for the experimental group using ChatGPT ($t(34) = -8.03$, $p < 0.001$, $d = 0.46$) aligns with a growing body of research affirming the utility of AI-based feedback in EFL instruction. As shown by Tseng and Lin [13], the implementation of GPT-3.5 with a rigid instructional design improved the quality of student writing in terms of parameters such as grammatical correctness, thematic consistency, and learner independence. On the same note, Alghannam [17] found that ChatGPT provided feedback was essentially correct and useful, particularly in addressing grammatical and structural issues. However, she observed that it was not always consistent. This can be used to illustrate the point that ChatGPT feedback is particularly suitable for lower levels of writing, such as A2, where the rudimentary aspects of sentence construction and grammar are typically the focus of learning.

In addition, Jacobsen and Weber [20] noted that the success of using LLM feedback was highly dependent on the specificity of the prompt. In this paper, the prompt used (Appendix B) was developed by the principal investigator to give structured, effective, and practical feedback. This selective style is likely the reason for the quality and reliability of the observed feedback. Combined with the results of Jacobsen and Weber [20], the results of this study suggest that, under proper guidance, ChatGPT has the potential to provide scalable, hands-on assistance for lower proficiency learners, challenging the notion that AI tools are only beneficial to highly skilled users.

The experimental group's significant improvement across language, content, and organization—especially in language (from 2.03 to 3.23)—mirrors findings by Liu and Ma [6], who observed lexical and grammatical gains through ChatGPT-assisted writing. Similarly, Polakova and Ivenz [9] reported enhanced clarity and structural precision when students engaged in iterative AI feedback. In contrast, Alghannam [17] noted that ChatGPT struggles to support content depth and critical thinking. This limitation was mitigated in our study through the researcher-designed prompt, "Evaluate this paragraph based on Cambridge A2-level standards (CEFR)…" (Appendix B), which focused ChatGPT's feedback on specific rubric criteria. These findings suggest that prompt-driven AI feedback can improve both linguistic accuracy and text structure, even at an A2 level.

Although Liu and Ma [6] reported positive results in terms of lexical richness and grammatical accuracy, as indicated by marks of AI-based feedback, these study findings contribute to the discussion by showing that structural accuracy also improves significantly at the A2 level. This dimension was not given much importance in the previous work. On the same note, Polakova and Ivenz [9] noted that students who received iterative AI feedback showed improvement in conciseness with surface-level precision, but reported that the ability to transfer any correction to higher-level text cohesion was challenged. Comparatively, the structured prompt we used efficiently reduced these issues and led to a higher level of uniformity in corrections across grammar, vocabulary, and organizational features. Therefore, this study has added by shedding light on how timely ChatGPT feedback can be used to promote aspects beyond lexical correction, such as textual cohesion and rhetorical structure.

Regarding cognition, the immediate response available with ChatGPT may help reduce the cognitive load associated with finding and fixing mistakes. As these low-level processes externalize, learners will be able to devote their

working memory resources to other higher-level processes, including planning, mental simulation, and argument development. The present finding aligns with earlier theories explicated in the study, which are associated with second language writing. These theories hold that working memory is a limited resource, serving as a source of developmental environment, and is limited by its capacity to handle syntactic complexity and coherence. In that regard, the feedback provided by ChatGPT is not limited to correcting superficial mistakes but also enables students, more indirectly, to learn how to address more challenging aspects of academic writing.

Remarkably, even the participants in the experimental group did not experience a loss of writing performance, indicating that incorporating ChatGPT did not pose a risk to learners' development. Such a result contrasts the apprehensions expressed by Teng and Huang [25], who warned that the unmoderated use of AI feedback can lead to cognitive disengagement and formulaic writing, especially in lower grades. In line with Huang [26], the study provided evidence that, when AI tools are integrated into a granular pedagogical structure--focused on active reflection, informed interpretation of the feedback, and iterative revisions--students can be cognitively alert and take advantage of the immediacy of replies provided in AI tools. This highlights the need to consider the necessity of building AI literacy as a precursor to sustainable integration, especially via outright instruction in prompt engineering and feedback assessment.

Nonetheless, in this study, a structured prompt and guided reflection helped promote learner independence and alleviate possible misuse. These results align with those of Kusuma et al. [12], who found that after training EFL preservice teachers in prompt engineering, they applied ChatGPT responsibly and beneficially as a tool to achieve their learning objectives. Furthermore, it has been projected that the repetitiveness of structured AI feedback could result in faster metacognitive control in writing among students who could easily apply the skills in their independent academic work without the support of AI [26,27]. Nevertheless, these advantages are only retainable through the recalibration of the prompts over time and through continuous mediation of the teacher to maintain the distance and ensure compliance with the changing curricular standards.

The outcomes, which indicated that the language scores within the control group did not improve (pre- and post-test averages of 2.20 points), point to the perennial grammatical stagnation that is common in conventional instruction. This finding aligns with Pasaribu et al. [1], who also identified syntactical and morphological errors, particularly errors with tenses, verb forms, and article usage, as a consistent issue among Indonesian EFL learners. Similarly, Mahmood [2] found that Pakistani undergraduates regularly struggled with grammatical control, reflecting a broader pedagogical gap in grammar instruction. In contrast, the experimental group in this study, which received ChatGPT feedback prompted a tailored command, showed marked language gains. This suggests that real-time, personified AI feedback has the potential to bridge those gaps by continually correcting and demonstrating appropriate usage. Nevertheless, as noted in prior literature, overreliance on automated corrective feedback can risk narrowing students' linguistic repertoire if not supplemented with tasks promoting creativity and higher-order thinking [26,27].

These constraints reflect previous research, which has found that AI feedback tends not to initiate higher-order processing and critical interaction [17]. However, in contrast to such reports, our intervention involved AI feedback and guided reflection, which reduced the risk of disengagement. This finding is partially in line with Teng and Huang [25], who warned that formulaic writing is a result of unscaffolding the use of AI. The present scaffold system, in turn, increases precision and autonomous control by the retainer. Thus, the research recommends that easy-to-administer prompts and reflective areas should be implemented to enhance the pedagogical value of AI tools, as well as make them more of a helpful aid to anxiety-inducing, in-depth aspects of study, rather than one that merely features a surface-level correct. An optimal integration model should balance AI-generated micro-level corrections with human-facilitated macro-level guidance on content development and rhetorical strategies.

Therefore, the study provides evidence that prompt-based AI support offers a scalable and pedagogically sound solution for addressing grammatical stagnation in beginner-level EFL contexts, while also highlighting the need for policy-level frameworks to guide ethical, equitable, and context-sensitive adoption of generative AI in formal education [26,27].

4.1. Limitations and Future Research

Although the study makes significant contributions, it is also limited by the fact that it only utilizes quantitative data. Although the statistical outcomes demonstrate performance gains, they do not capture the subjective experience of learners, including their perceptions and the cognitive models that influence their revisions. Such a lack of qualitative data limits our understanding of how learners perceive and respond to AI feedback. In future interviews, reflective journals or think-aloud protocols may be included to provide a comprehensive view of the learner-AI interaction. Additionally, the findings lack generalizability due to the small sample size and the study's limited scope, which only covers one institution in Ecuador.

A limitation of this study is that the teacher who provided feedback to the control group was not blinded to the group assignment, which may have inadvertently affected the style, frequency, or depth of feedback. Nonetheless, group tasks, weekly schedule, and lesson planning were strictly observed in both groups, which implied that teaching delivery was not purposely changed. In principle, this minimized the risk of bias without eliminating it.

In addition, the use of a standardized rubric and independent double-rating of pre- and post-test writing samples—supported by acceptable inter-rater reliability values—further reduces the likelihood that this potential bias substantially altered the outcome. These procedures provide a sense of assurance that the stated changes in the experimental group can be explained primarily by the structured ChatGPT feedback, rather than the inconsistency in instructional delivery.

This limitation can be overcome in future studies by using blinded evaluators or multiple instructors to reduce the possibility that teacher knowledge will affect student performance.

Moreover, future research should extend this inquiry to other CEFR proficiency levels (e.g., B1, B2) to explore whether similar improvements are observed in more complex writing tasks. Studies comparing AI-generated feedback with either teacher- or peer-written feedback would also shed some interesting light on pedagogy trade-offs, especially concerning the quality, depth, and learning outcomes of students. Long-term research could also determine the lasting effect of AI implementation on writing progress after several semesters.

5. Conclusions

In conclusion, this study investigated the effectiveness of ChatGPT-generated feedback in improving academic writing skills among A2-level EFL learners. The results indicated that learners exposed to AI received more structured feedback, with significant variations in writing skills, including language accuracy, coherence, and organization, compared to learners in the control groups who received feedback using original techniques.

These results validate that generative AI, when guided by pedagogically designed prompts, can function as a reliable supplementary feedback mechanism even for low-proficiency learners. The absence of any performance decline among experimental participants also confirms that structured AI use did not produce cognitive disengagement or mechanical writing patterns, contrary to concerns raised by Teng and Huang [25]. Instead, it supported learner autonomy and improvement through iterative, rubric-based evaluation.

The structured application of prompts, constructed by the leading researcher to meet the requirements of the Cambridge A2-level, proved to be a vital element in providing focused, rubric-aligned feedback. The results align with the earlier assertions [6,20] regarding the effectiveness of the prompts used in generating valuable feedback from AI. Generally, the research demonstrates that when properly used, ChatGPT can assist in filling the instructional gaps related to A2-level scholars' writing, particularly in grammar and text structure, and it does not pose a threat to learners' development.

By assigning students a brief writing task and submitting it to ChatGPT, teachers can incorporate structured instructions into weekly classes and explain how to interpret the comments generated by AI. This strategy proved successful over the 12 weeks of this study. One example of how integration could be carried out is to devote the last 15 minutes of each writing session to analyzing the comments according to the instructions. This would encourage students to compare the suggestions provided by the AI with those of the teacher and to think critically about the credibility and effectiveness of the comments generated by ChatGPT. This methodology would not only allow the intervention to be replicated. However, it would also provide educators with clear guidelines on how to incorporate AI comments based on instructions into current lesson plans without disrupting

the curriculum.

Given the promising results, it is recommended that EFL teachers incorporate AI tools, such as ChatGPT, into their writing instruction, especially for beginner students who require immediate, repetitive, and individualized feedback. Nevertheless, to ensure the responsible use of AI and develop AI literacy in students, educators should conduct training on quick formulation and feedback interpretation [16]. Teacher mediation is also critical for maintaining a balance between interpreting feedback and avoiding dependence on AI suggestions.

Author Contributions

Conceptualization, J.I.H.V.; methodology, J.I.H.V.; validation, J.I.H.V., J.C.d.S., and R.M.V.B.; formal analysis, J.I.H.V., J.C.d.S., and R.M.V.B.; investigation, J.I.H.V.; resources, J.I.H.V.; data curation, J.I.H.V.; writing—original draft preparation, J.I.H.V.; writing—review and editing, J.C.d.S., and R.M.V.B.; visualization, J.I.H.V.; supervision, J.C.d.S.; project administration, J.I.H.V. and J.C.d.S. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki and approved by the Coordination Office of the Pedagogy of National and Foreign Languages Program at Quevedo State Technical University (Approval No. 03, dated December 5th, 2024), which granted formal permission to conduct the research after reviewing a formal letter describing the study's objectives, methodology, and procedures.

Informed Consent Statement

All participants signed a written informed consent form prior to the commencement of data collection. All participants were over 18 years of age and did not require authorization from their legal guardians. Before enrolling, they were informed about the study objectives, procedures, potential risks and benefits, and confidentiality measures in an induction talk. Participation was entirely voluntary, and students were informed of their right to withdraw at any time without penalty or consequence. The data were anonymized to protect the participants' identities.

Data Availability Statement

The dataset generated and analyzed during the current study is publicly available in the Figshare repository at https://doi.org/10.6084/m9.figshare.29755598.v1. All data are anonymized and prepared for replicable use. The research instruments, including the ChatGPT prompt and the rubric, are provided in the appendices of this manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix A

Table A1. Writing Assessment Subscales for A2 Key for Schools.

A2	Content	Organisation	Language
5	All content is relevant to the task. Target reader is fully informed.	Text is connected and coherent, using basic linking words and a limited number of cohesive devices.	Uses everyday vocabulary generally appropriately, while occasionally overusing certain lexis. Uses simple grammatical forms with a good degree of control. While errors are noticeable, meaning can still be determined.
4	*Performance shares features of Bands 3 and 5.*		
3	Minor irrelevances and/or omissions may be present. Target reader is on the whole informed.	Text is connected using basic, high-frequency linking words.	Uses basic vocabulary reasonably appropriately. Uses simple grammatical forms with some degree of control. Errors may impede meaning at times.

Table A1. *Cont.*

A2	Content	Organisation	Language
2	*Performance shares features of Bands 1 and 3.*		
1	Irrelevances and misinterpretation of task may be present. Target reader is minimally informed.	Production unlikely to be connected, though punctuation and simple connectors (i.e. 'and') may on occasion be used.	Produces basic vocabulary of isolated words and phrases. Produces few simple grammatical forms with only limited control.
0	Content is totally irrelevant. Target reader is not informed.	*Performance below Band 1.*	

Appendix B. ChatGPT Command

"Evaluate this paragraph based on Cambridge A2-level standards (CEFR) for Content, Organization, and Language (vocabulary and grammar). Provide specific feedback."

References

[1] Pasaribu, A.N., Pasaribu, T.K., Siahaan, R.B., et al., 2024. Challenges for Indonesian EFL students in English academic writing. English Review: Journal of English Education. 12(2), 527–536. DOI: https://doi.org/10.25134/erjee.v12i2.9780

[2] Mahmood, K., 2021. Academic writing challenges of EFL learners and teachers' proficiency in Pakistani higher education. Journal of Research in Social Sciences. 8(2), 56–57. DOI: https://doi.org/10.52015/jrss.8i2.55

[3] 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, 1260843. DOI: https://doi.org/10.3389/fpsyg.2023.1260843

[4] Alostath, K., 2021. Teachers' perception of EFL students' poor writing skills: The challenges, causes, and remedies. International Journal of Science Research (Raipur). 10(4), 647–651. DOI: https://doi.org/10.21275/sr21413012657

[5] Zhou, T., Cao, S., Zhou, S., et al., 2023. Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. System. 118, 103141. DOI: https://doi.org/10.1016/j.system.2023.103141

[6] Liu, G., Ma, C., 2024. Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. Innovation in Language Learning and Teaching. 18(2), 125–138. DOI: https://doi.org/10.1080/17501229.2023.2240316

[7] Ibrahim, K., Kirkpatrick, R., 2024. Potentials and implications of ChatGPT for ESL writing instruction. International Review of Research in Open and Distributed Learning. 25(3), 394–409. DOI: https://doi.org/10.19173/irrodl.v25i3.7820

[8] Krause, S., Panchal, B.H., Ubhe, N., 2025. The evolution of learning: Assessing the transformative impact of generative AI on higher education. In: Lecture Notes on Data Engineering and Communications Technologies. Springer Nature: Singapore. pp. 356–371. DOI: https://doi.org/10.1007/s44366-025-0058-7

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

[10] Ng, D.T.K., Tan, C.W., Leung, J.K.L., 2024. Empowering student self-regulated learning and science education through ChatGPT: A pioneering pilot study. British Journal of Educational Technology. 55(4), 1328–1353. DOI: https://doi.org/10.1111/bjet.13454

[11] Theophilou, E., Koyutürk, C., Yavari, M., et al., 2023. Learning to prompt in the classroom to understand AI limits: A pilot study. In Proceedings of the AIxIA 2023 – Advances in Artificial Intelligence: XXIInd International Conference of the Italian Association for Artificial Intelligence, AIxIA 2023, Rome, Italy, 6–9 November 2023. DOI: https://doi.org/10.1007/978-3-031-47546-7_33

[12] Kusuma, I., Roni, M., Dewi, K., et al., 2024. Revealing the potential of ChatGPT for English language teaching: EFL preservice teachers' teaching practicum experience. Studies in English Language and Education. 11(2), 650–670. DOI: https://doi.org/10.24815/siele.v11i2.34748

[13] Tseng, Y.C., Lin, Y.H., 2024. Enhancing English as a Foreign Language (EFL) learners' writing with ChatGPT: A university-level course design. Electronic Journal of E-Learning. 22(2), 78–97. DOI: https://doi.org/10.34190/ejel.21.5.3329

[14] Xiao, F., Zhu, S., Xin, W., 2025. Exploring the landscape of generative AI (ChatGPT)-powered writing instruction in English as a Foreign Language education: A scoping review. ECNU Review of Education. 1–19. DOI: https://doi.org/10.1177/20965311241310881

[15] Li, J., Huang, J., Wu, W., et al., 2024. Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. Humanities and Social Sciences Communications. 11(1). DOI: https://doi.org/10.1057/s41599-024-03755-2

[16] Nguyen, A., Duong, A.T., Nguyen, D.T.B., et al.,

2025. Guidelines for learning design and assessment for generative artificial intelligence-integrated education: A unified view. Information and Learning Sciences. ahead-of-print. DOI: https://doi.org/10.1108/ils-11-2024-0148

[17] Alghannam, M.S.M., 2024. Artificial intelligence as a provider of feedback on EFL student compositions. World Journal of English Language. 15(2), 161. DOI: https://doi.org/10.5430/wjel.v15n2p161

[18] Duc, A.N., Lønnestad, T., Sundbø, I., et al., 2023. Generative AI in undergraduate information technology education – Insights from nine courses. arXiv. arXiv:2311.10199. DOI: https://doi.org/10.48550/arXiv.2311.10199

[19] Tesch, A., 2016. Implementing pre-post test designs in higher education evaluations. New Directions for Evaluation. 2016(151), 85–96. DOI: https://doi.org/10.1002/ev.20195

[20] Jacobsen, L., Weber, K., 2025. The promises and pitfalls of LLMs as feedback providers: A study of prompt engineering and the quality of AI-driven feedback. OSF Preprints. DOI: https://doi.org/10.31219/osf.io/cr257

[21] Cambridge Assessment English, 2024. Teacher guide for writing A2 Key for Schools. Available from: https://www.cambridgeenglish.org/Images/600973-teacher-guide-for-writing-a2-key-for-schools.pdf (cited 5 December 2024).

[22] Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics. 33(1), 159–174. DOI: https://doi.org/10.2307/2529310

[23] Field, A., 2024. Discovering Statistics Using IBM SPSS Statistics, 6th ed. SAGE Publications: London, UK.

[24] Hoyos Valdiviezo, I., 2025. Dataset: ChatGPT feedback interventions for A2-level EFL academic writing. Available from: https://doi.org/10.6084/m9.figshare.29755598.v1 (cited 10 August 2025).

[25] Teng, M.F., Huang, J., 2025. Incorporating ChatGPT for EFL writing and its effects on writing engagement. International Journal of Computer-Assisted Language Learning and Teaching. 15(1), 1–21. DOI: https://doi.org/10.4018/ijcallt.367874

[26] Huang, J., 2023. Engineering ChatGPT prompts for EFL writing classes. International Journal of TESOL Studies. 5(4), 73–79. DOI: https://doi.org/10.58304/ijts.20230405

[27] Gozali, I., Wijaya, A.R.T., Lie, A., et al., 2024. Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. JALT CALL Journal. 20(1), 1–22. DOI: https://doi.org/10.29140/jaltcall.v20n1.1200