Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), marzo-abril 2026,
Volumen 10, Número 2.

https://doi.org/10.37811/cl_rcm.v10i2

IMPACT OF CORRECTIVE METALINGUISTIC WRITTEN FEEDBACK GIVEN THROUGH CHATGPT ON EFL PRE-SERVICE TEACHERS' ACADEMIC WRITING SKILLS

EL IMPACTO DE LA RETROALIMENTACIÓN METALINGÜÍSTICA ESCRITA CORRECTIVA PROPORCIONADA A TRAVÉS DE CHATGPT EN LAS HABILIDADES DE ESCRITURA ACADÉMICA DE DOCENTES EN FORMACIÓN DE INGLÉS COMO LENGUA EXTRANJERA.

Stewart Smith Viáfara Pérez
Universidad El Bosque

Daniela Sandoval Lasso
Universidad El Bosque

DOI: https://doi.org/10.37811/cl_rcm.v10i2.23809

Impact of corrective metalinguistic written feedback given through ChatGPT on EFL pre-service teachers' academic writing skills

Stewart Smith Viáfara Pérez[1]
sviafarap@unbosque.edu.co
https://orcid.org/0009-0005-1303-4167
Universidad El Bosque
Colombia

Daniela Sandoval Lasso
dsandovalla@unbosque.edu.co
https://orcid.org/0009-0003-9118-5345
Universidad El Bosque
Colombia

ABSTRACT

This study investigates the impact of ChatGPT-generated corrective metalinguistic written feedback on the academic writing performance of EFL undergraduate pre-service teachers. We implemented a true experimental pretest-posttest control group design with 14 participants randomly allocated to an experimental group (ChatGPT feedback) and a control group (teacher feedback). In addition, an analytic rubric that rated Grammar and Accuracy, Vocabulary Range and Appropriacy, Cohesion and Coherence, Task Achievement, Organization' perception questionnaires and open-ended replies complemented the quantitative data to evaluate writing performance. Descriptive data indicated enhancement in both groups, with the experimental exhibiting superior mean gain scores ($M = 7.29$) compared to the control group ($M = 6.29$). Paired-samples t-tests demonstrated statistically significant improvements from pre-test to post-test in both groups ($p < .001$). Category-level analysis indicated stronger gains for the experimental group in Task Achievement, Organization, Grammar, and Accuracy. Perception results indicated more positive assessments of clarity and actionability in the experimental group (e.g., 71% strongly concurred that the feedback was clear, and comprehensible). Findings suggest that ChatGPT-mediated metalinguistic feedback can effectively enhance academic writing growth and may function as an auxiliary feedback resource in EFL teacher education settings.

Keywords: Metalinguistic Written Feedback; ChatGPT; Corrective feedback; EFL Writing Skills

[1] Autor principal
Correspondencia: sviafarap@unbosque.edu.co

El impacto de la retroalimentación metalingüística escrita correctiva proporcionada a través de ChatGPT en las habilidades de escritura académica de docentes en formación de inglés como lengua extranjera.

RESUMEN

Este estudio investiga el impacto de la retroalimentación escrita metalingüística correctiva generada por ChatGPT en el rendimiento de la escritura académica de los futuros profesores de inglés como lengua extranjera (EFL) a nivel de pregrado. Implementamos un diseño de grupo control pretest-post-test experimental verdadero con 14 participantes asignados aleatoriamente a un grupo experimental (retroalimentación de ChatGPT) y un grupo control (retroalimentación del profesor). Asimismo, la evaluación del desempeño escrito utilizó una rúbrica analítica que calificó la gramática y la precisión, la amplitud y la adecuación del vocabulario, la cohesión y la coherencia, así como la consecución de la tarea y la organización. Los cuestionarios de percepción y las respuestas abiertas complementaron los datos cuantitativos. Los datos descriptivos indicaron una mejora en ambos grupos, con el experimental mostrando puntuaciones medias de ganancia superiores (M = 7.29) en comparación con el grupo de control (M = 6.29). Las pruebas t de dos muestras indicaron mejoras estadísticamente significativas desde el pre-test al post-test en ambos grupos ($p < .001$). El análisis de la rúbrica por categorías revelo ganancias más fuertes para el grupo experimental en el cumplimiento la tarea, Organización, Gramática y Precisión. Los resultados de la percepción de cada uno de los participantes indicaron evaluaciones más positivas de claridad y viabilidad en el grupo experimental (por ejemplo, el 71% estuvo totalmente de acuerdo en que la retroalimentación era clara y comprensible). Los resultados sugieren que la retroalimentación metalingüística generada a través de ChatGPT puede mejorar eficazmente el crecimiento de la escritura académica y a su vez podría funcionar como un recurso de retroalimentación auxiliar en entornos de formación de profesores de inglés como lengua extranjera.

Palabras clave: Retroalimentación metalingüística escrita; ChatGPT; Retroalimentación correctiva; EFL Habilidades de escritura

Artículo recibido 02 abril 2026
Aceptado para publicación: 30 abril 2026

INTRODUCTION

In the past decades, the integration of artificial intelligence (AI) has come to play a significant role in the educational context, especially in the fields of applied linguistics and language education. Writing, as one of the most complex and demanding skills in second language acquisition, has long posed challenges for learners and educators alike (Hyland, 2019). Corrective feedback, particularly in written form, is seen as a feasible method to address these challenges. Bitchener and Ferris (2012) argue that this type of feedback has been widely recognized as a pivotal pedagogical tool for enhancing learners' accuracy and fluency. Among the various feedback mechanisms, metalinguistic corrective feedback, by which the nature of an error is explained explicitly, has been shown to promote noticing and long-term language development (Ellis, 2009; Lyster, Saito, & Sato, 2013). For the most part, such feedback has been provided by teachers; nonetheless, the increasing demand for individualized and immediate feedback has been a key factor in exploring alternative technologies.

With the emergence of AI tools such as ChatGPT, the alternatives for providing immediate and personalized corrective feedback have increased significantly. ChatGPT has the potential to generate metalinguistic explanations and suggest revisions in real time (Barrot, 2023), which may foster learner agency beyond the instructional context. Corrective feedback has been studied in second language writing; nevertheless, empirical evidence regarding the impact of ChatGPT's corrective metalinguistic feedback on pre-service teacher's writing development remains limited. Existing studies have essentially focused on teacher and peer feedback, leaving a gap in how AI-generated metalinguistic feedback influences learners' writing accuracy.

With that in mind, this research aimed to address this gap by examining the impact of corrective metalinguistic written feedback generated by ChatGPT in the writing skills of undergraduate students in bilingualism (English and Spanish) in a Colombian university. We carried out this study considering both the results of the students' argumentative essays and their reflections on the writing process, including whether they perceived any improvement in their writing skills. Additionally, by adopting a mixed-method approach, the study made a distinction between learners who received ChatGPT-based feedback and those who received traditional teacher feedback across pre and posttests essay writing tasks. By doing so, this research aimed to offer a comprehensible understanding of the impact of using

this AI in written tasks, enabling teachers to incorporate ChatGPT into their feedback practices in a safe and reliable manner, thereby facilitating and streamlining classroom instruction, as well as fostering autonomy in the students' learning process, empowering them to actively participate in their own educational journey.

LITERATURE REVIEW

Academic Writing

Academic writing is widely regarded as one of the most demanding skills in EFL learners, as it requires both linguistic accuracy and mastery of discourse organization and genre conventions (Hyland, 2019). In higher education, argumentative essays represent a crucial genre by which students demonstrate critical thinking, adopting a position, and coherence in developing ideas (Wingate, 2012). Undergraduate students in language related programs face pressure to achieve this proficiency, as their academic and future professional performance depends on their ability to write with clarity, accuracy, and persuasiveness. Research has shown that analytic rubric which distinguishes between grammar, vocabulary, cohesion, and organization provides a reliable way to capture the multifaceted nature of writing proficiency (Weigle, 2002; Andrade, 2005). This multifaceted perspective emphasizes the necessity for focused instructional and feedback methodologies that may tackle both form and content.

Corrective and metalinguistic written feedback

Corrective written feedback (WCF) has been recognized as a central pedagogical tool to support such development. Ferris (2010) and Bitchener and Ferris (2012) highlight its role in drawing learners' attention to linguistic errors while enhancing long-term accuracy and fluency. Within this framework, metalinguistic corrective feedback, feedback that not only identifies an error but also explains the underlying rule or principle, has been found to enhance learners' noticing and promote greater self-regulation in revision (Ellis, 2009; Lyster, Saito, & Sato, 2013). Previous studies suggest that metalinguistic feedback is particularly effective in academic contexts, where learners are expected to internalize conventions and apply them autonomously across genres (Shintani & Ellis, 2015). Therefore, grounding corrective practices in explicit explanation rather than simple error indication aligns appropriately with the needs of undergraduate EFL students, especially those preparing to become language professionals.

AI-mediated feedback in language learning

The development of artificial intelligence (AI) tools in education has introduced new possibilities within the field of language acquisition, especially in writing feedback. Recent studies mention that AI-driven platforms such as ChatGPT can provide immediate, individualized, and linguistically rich feedback that mirrors human corrective responses (Zou et al., 2023a). In contrast to automated grammar checkers, ChatGPT can generate explanations, examples, and alternative phrasing, which align closely with the principles of metalinguistic feedback. This immediacy and adaptability may enhance learners' engagement and autonomy by allowing students to explore error and corrections in real time (Kasneci et al., 2023). Nonetheless, concerns remain about accuracy and overreliance. Likewise, the lack of studies addressing its role in providing corrective metalinguistic written feedback in argumentative essays represents an important research gap. Although the pedagogical applications of ChatGPT are still in their early stages, initial empirical findings show its capability to complement or even extend traditional forms of teacher feedback. EFL learners perceived ChatGPT's feedback as clear and useful for enhancing written summaries, while at the same time, teacher guidance remained essential in understanding complex explanations (Poláková and Ivenz, 2024). Similarly, Lin (2023) states that ChatGPT can support learners' academic writing development effectively, especially in structuring arguments and expanding lexical variety. Nonetheless, concerns remain about accuracy and overreliance. Additionally, the lack of studies addressing its role in providing corrective metalinguistic written feedback in argumentative essays represents an important research gap.

Therefore, this suggests that ChatGPT may serve as a valuable supplementary tool in higher education contexts, especially for undergraduate EFL students who must master academic writing. Exploring this intersection is critical for understanding whether AI-generated feedback can foster not only linguistic accuracy but also higher order writing skills such as cohesion, organization, and critical argumentation.

RESEARCH QUESTIONS

This study addresses the following research questions:

1. To what extent does corrective metalinguistic written feedback provided by ChatGPT impact pre-service EFL teachers' argumentative writing performance compared to traditional teacher feedback?

2. How do students perceive the usefulness of ChatGPT feedback in relation to teacher feedback?

OBJECTIVES

GENERAL OBJECTIVE

• To examine the effectiveness of ChatGPT-generated corrective metalinguistic written feedback compared to teacher feedback on EFL pre-service teachers' academic writing performance.

SPECIFIC OBJECTIVES

• To compare pretest and posttest writing scores between students receiving ChatGPT feedback and those receiving teacher feedback.

• To analyze students' perceptions of the feedback process through questionnaires

METHODS

This research is framed within a true experimental pretest-posttest control-group design, with participants randomly assigned to either the experimental group (here after EG), which received ChatGPT-generated corrective metalinguistic written feedback, or the control group (CG), where the feedback was provided by the teacher. This design ensures equivalence between groups at the start of the intervention, allowing stronger causal inferences regarding the effectiveness of the feedback type on academic writing performance (Campbell & Stanley, 1963; Cohen et al., 2018). In addition to this, the present study studied adopted a mixed-methods approach, integrating quantitative instruments (essay rubrics, background and post-questionnaires) with qualitative insights (perception questionnaire and student reflections) to provide a comprehensive understanding of the intervention's effect (Creswell & Creswell, 2018).

PARTICIPANTS

A total of 14 undergraduate students were part of this study, aged between 19 and 23 years old; enrolled in the Bachelor of Foreign Languages with an Emphasis on English program at a private university in Bogotá, Colombia. All participants reported previous experience using ChatGPT, both in academic and work-related contexts. In accordance with the study design, we randomly allocated students to two groups: a control group that received teacher feedback and an experimental group that received ChatGPT-generated feedback (Lammers & Badia, 2011), which received ChatGPT-generated feedback. The independent variable was the source of feedback (ChatGPT vs teacher), while the dependent

variables were academic performance, assessed through analytic rubrics, and students' perceptions, gathered via qualitative questionnaires (Panjaitan, 2018).

INSTRUMENTS

As Barrot (2023) highlights, ChatGPT has recently been recognized for its potential to enhance feedback processes in second-language education by providing learners with immediate, personalized, and metalinguistic input. Accordingly, the core instrument in this study was ChatGPT, which served as the feedback tool for the EG by providing corrective metalinguistic written feedback on pre-service teachers' essays. Its output was standardized through carefully designed prompts to ensure consistent coverage of grammar, vocabulary, cohesion, and task achievement. Simultaneously, the CG received feedback from the course professor using the same categories, enabling direct comparison. Additional instruments included (A) argumentative writing tasks, used for pretest and posttest essays (Alarcon & Morales, 2011). (B) a scoring rubric applied by two independent raters to assess writing performance across four dimensions (Brookhart & Chen, 2015). (C) questionnaires (background and post-questionnaire) to gather demographic information, expectations, and perceptions of the feedback received (Aksu, 2009); and (D) a Revision Log sheet in which undergraduate students recorded how they responded to specific feedback points (Yi et al., 2014). Together, these instruments provided both quantitative and qualitative data for analysis.

PROCEDURE

The intervention spanned six weeks and focused on a single major argumentative writing task (A), implemented through four key instructional stages comprising drafting, feedback, and revision. During the first stage, participants completed a short questionnaire about their prior use and experience with ChatGPT, along with a diagnostic task to identify participants' prior knowledge of an argumentative essay structure. We gave the whole group an essay prompt: "Do you think animals can really communicate with humans in meaningful ways?" The instruction was to produce a 240-word argumentative essay without using any AI assistance. The purpose of this intervention was to know the participants' familiarity with the Chat GPT tool and evaluate prior knowledge of essay writing.

Following this, during the second stage, students were familiarized with the structure of an academic outline, including an introduction (a hook, background information, and a thesis statement), body

paragraphs, and a conclusion. This structure was incorporated to guarantee consistency across participants and to enhance explicit genre awareness as suggested by Hyland (2019) and Ferris (2020) in L2 writing pedagogy. This stage was essential to identify the differences between the initial diagnostic task and the subsequent writing performance.

In the third stage, participants were asked to begin the second writing task (Task A), applying the previously explained structure and responding to the same essay prompt, thereby marking the onset of the treatment phase of the study. All students completed this task without using any AI assistance.

Participants in the experimental group (EG) received ChatGPT-generated corrective metalinguistic written feedback on their drafts through a standardized research design prompt. This prompt guided ChatGPT to identify and explain up to fifteen language related issues, categorized under Grammar, Vocabulary, Cohesion, and Organization, providing rule-based metalinguistic explanations and examples for self-revision. Conversely, the control Group (CG) received teacher feedback by covering the same four analytic dimensions to maintain consistency in focus and scope. This procedure aligned with previous studies comparing automated and teacher feedback in ELF writing (Li & Li, 2023; Poláková & Ivenz, 2024).

Once the feedback was handed out, we had both groups dive into their revisions and fill out a Revision Log. The idea was to track their initial errors against the type of feedback they got and the specific changes they ended up making. It was meant to spark some real reflection and to develop self-regulated learning. Plus, these logs gave us a clear window into how they were using those metalinguistic explanations when they sat down to polish their drafts.

Finally, during the last stage, participants completed the post-test, which involved writing a new essay on the same prompt under comparable classroom conditions and without any external feedback. This stage enabled us to assess improvement in overall writing proficiency as well as performance across each of the five (5) different categories. Afterwards, students completed a post-questionnaire designed to collect data on their perceptions of the feedback experience, their confidence in academic writing, and their perceived progress.

We analyzed all collected data, including essays, feedback samples, revision logs, and questionnaires, quantitatively and qualitatively to determine the effectiveness of ChatGPT generated metalinguistic feedback in enhancing pre-service EFL teachers' academic writing skills.

RESULTS AND FINDINGS

This section elaborates on the results of the pre- and post-tests that were conducted to evaluate the effectiveness of ChatGPT-generated corrective metalinguistic written feedback compared with the traditional teacher feedback on EFL pre-service teacher's academic writing performance. The analysis focuses on participants' improvement in overall writing quality along with the respective descriptive and inferential statistics comparing the experimental and control groups.

Table 1 presents the descriptive statistics for both groups. Overall, both the CG and the EG showed improvement from pre-test to post-test. The EG started with a slightly higher pre-test mean (M = 9.71, SD = 1.80) and also achieved a notably higher post-test mean (M = 17.00, SD = 1.00) compared to the CG (M = 8.86, SD = 1.68; M = 15.14, SD = 2.73). In line with this pattern, the mean gain score was greater for the EG (M = 7.29, SD = 2.14) than for the CG (M = 6.29, SD = 1.80), suggesting stronger overall progress among participants who received ChatGPT-mediated metalinguistic written feedback.

Paired-samples t-tests confirmed that these improvements were statistically significant in both groups. EG showed a substantial increase from the pre-test to the post-test ($t(6) = 9.02$, $p < .001$, Cohen's dz = 3.41), indicating a large effect size. Similarly, the CG also exhibited a significant improvement ($t(6) = 9.24$, $p < .001$, Cohen's dz = 3.49), reflecting a comparably large effect.

Despite substantial advancements by both groups, the differences in overall improvement between ChatGPT and teacher feedback was not statistically significant in this sample, as indicated by an independent-samples t-test comparing gain scores between groups ($t(12) = 0.95$, $p > .05$).

Category-level Analysis

Four analytical categories - Grammar & Accuracy, Vocabulary Range & Appropriateness, Cohesion & Coherence, and Task Achievement & Organization - were employed to evaluate mean scores to examine enhancements in specific writing components. Table 2 illustrates that both groups achieved progress. Nevertheless, the EG consistently demonstrated slightly greater gains. The most significant improvement for this group was observed in Task Achievement & Organization (+2.00), followed by

Cohesion & Coherence (+1.86) and Grammar & Accuracy (+1.86). Similarly, the CG exhibited comparable enhancements, particularly in Cohesion & Coherence (+1.71) and Task Achievement & Organization (+1.86). Overall, these findings indicate that both feedback conditions supported develpment at both the micro level (e.g., grammar and vocabulary) and macro level (e.g., organization and cohesion) with a slight descriptive advantage for the ChatGPT condition.

Perception Data and Qualitative Insights

As shown in Chart 1, the quantitative perception data indicated more favorable evaluations in the experimental group. Seventy-one percent of participants in the EG strongly agreed that ChatGPT's feedback was clear and comprehensible, compared to1 4% of participants in the CG regarding the clarity of teacher comments. In addition, 85% of EG participants reported that ChatGPT facilitated meaningful revision, whereas 57% of CG participants strongly agreed that teacher feedback led to noticeable changes in their writing. Self-reported confidence levels increased in both groups across all four writing domains, with slightly higher post-intervention in their writing.

Light thematic analysis of open-ended responses further supported these trends. Participants in the experimental group underscored the significance of clear metalinguistic elucidations and structural direction. One participant remarked that ChatGPT elucidated "verb tenses and correct word order" (P2, EG), while another emphasized feedback on "comma splice" problems (P12, EG). A number of students additionally indicated heightened knowledge of essay structure, particularly the significance of a "strong hook" (P10, EG).

Conversely, participants in the CG focused primarily on improving organization and building confidence. Students reported greater ability to use connectors such as "however" and "for example" (P7, CG) as well as to structure their ideas more coherently (P6, CG). Some participants highlighted the motivational value of positive instructor feedback (P5, CG).

Overall, the qualitative data suggest that participants perceived ChatGPT feedback as explicit and rule-oriented, whereas instructor feedback was associated with structural guidance and interpersonal support (See Chart 2).

Illustrations, tables and figures

Table 1.
Descriptive statistics (Pre- and Post-Test writing scores by Group)

Group	Pre-Test *M* (SD)	Post-Test *M* (SD)	Gain *M* (SD)	Min-Max Gain
CG	8.86 (1.68)	15.14 (2.73)	6.29 (1.80)	4 - 8
EG	9.71 (1.80)	17.00 (1.00)	7.29 (2.14)	4 -10

Note. CG = Control Group. EG = Experimental Group. *M* = Mean. *SD* = Standard deviation.

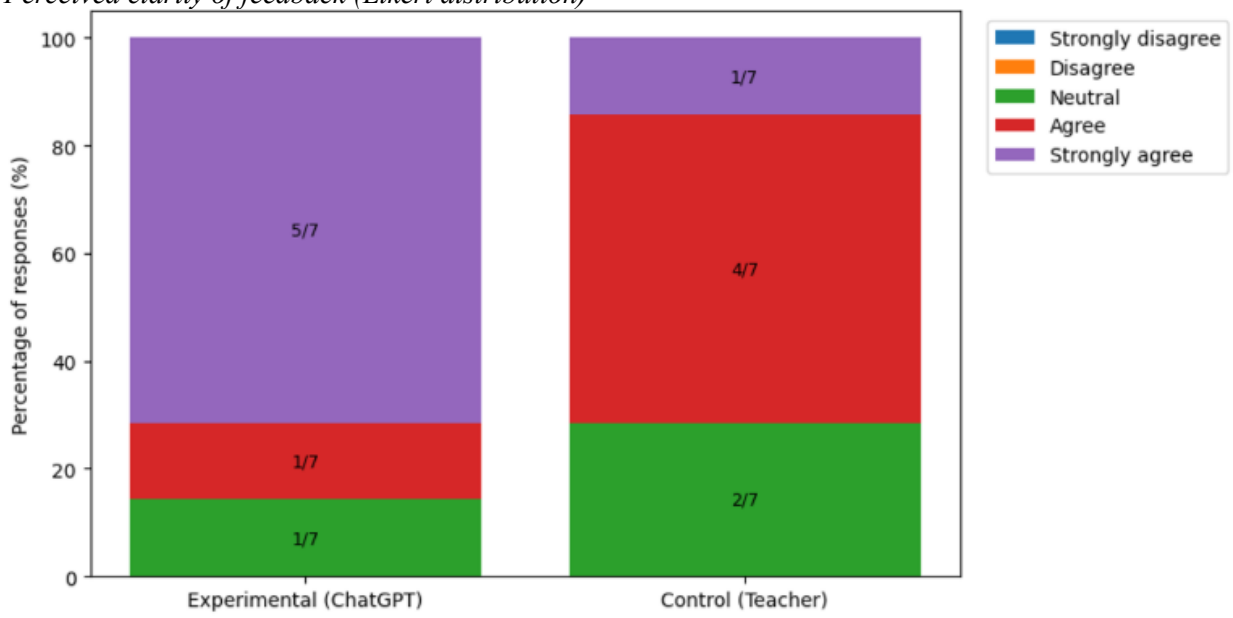
Table 2.
Category-Level Descriptive Statistics (Pre- and Post-Test by Group)

Category	Group	Pre-Test *M* (SD)	Post-Test *M* (SD)	Mean Gain
Grammar & Accuracy	CG	1.86 (0.38)	3.43 (0.79)	+1.57
	EG	2.14 (0.90)	4.00 (0.00)	+1.86
Vocabulary Range & Appropriacy	CG	2.29 (0.49)	3.29 (0.76)	+1.00
	EG	2.29 (0.49)	3.86 (0.38)	+1.57
Cohesion & Coherence	CG	2.29 (0.76)	4.00 (0.58)	+1.71
	EG	2.43 (0.53)	4.29 (0.49)	+1.86
Task Achievement & Organization	CG	2.43 (0.53)	4.29 (0.76)	+1.86
	EG	2.86 (0.69)	4.86 (0.38)	+2.00

Note. CG = Control Group. EG = Experimental Group. *M* = Mean. *SD* = Standard deviation.

Chart 1.

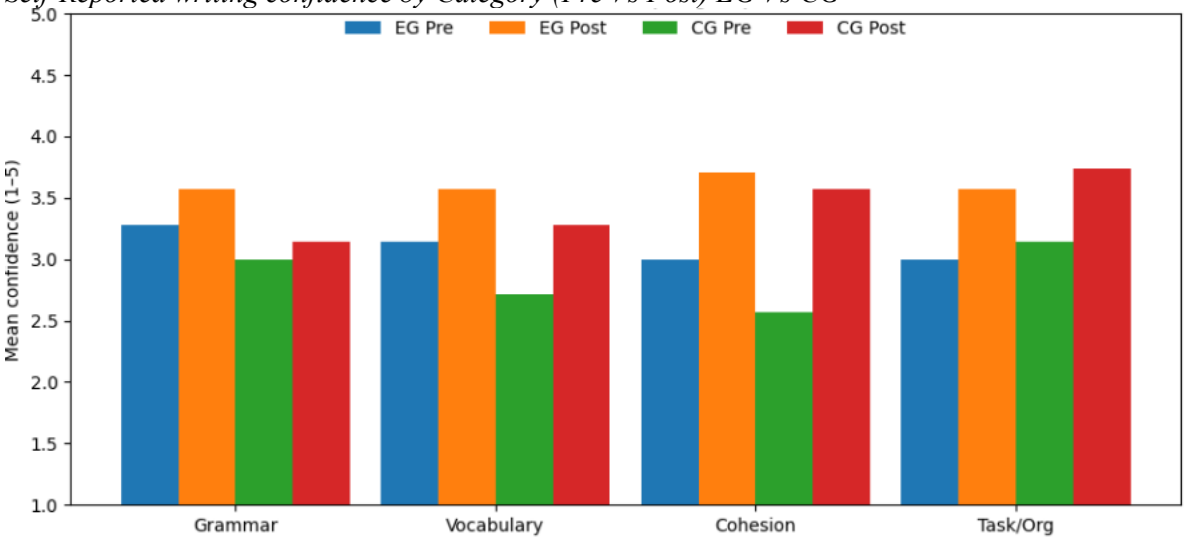
Perceived clarity of feedback (Likert distribution)



Note. $n = 7$ per group

Chart 2.

Self-Reported writing confidence by Category (Pre Vs Post) EG Vs CG



Note. $n = 7$ per group

DISCUSSION

This study examined the impact of ChatGPT-generated corrective metalinguistic written feedback on the academic writing performance of undergraduate EFL pre-service teachers, in contrast to traditional teacher feedback. The results showed substantial improvement from pre-test to post-test in both groups,

with statistically significant gains in each condition. While the EG demonstrated slightly greater descriptive gains and more positive perceptions of feedback, clarity, and usefulness, the differences in gain scores between the groups were not statistically significant. Overall, these findings suggest that ChatGPT-mediated metalinguistic feedback can effectively support the development of academic writing and may serve as a valuable complement to instructor feedback in EFL pre-service teacher education.

A key takeaway from the findings is that both feedback modes were associated with significant improvements in writing performance. This corresponds with known studies demonstrating that written corrective feedback can enhance awareness and foster improvement in L2 writing, especially when learners are compelled to modify and contemplate their draft (Bitchener & Ferris, 2012). The substantial within-group effect sizes notes in both scenarios indicate that organized feedback, coupled with revision opportunities, can provide significant improvements even within a brief instructional timeframe. This sheds light on RQ1, as the results demonstrate that metalinguistic feedback derived from ChatGPT yielded learning improvements similar to those achieved through teacher input. Although the EG group exhibiting somewhat superior descriptive gains and elevated post-test averages, the intergroup gain of disparity was not statistically significant.

At the category level, both groups demonstrated improvements across all analytic domains, with particularly notable gains in Task Achievement and Organization, as well as in Cohesion and Coherence. These patterns suggest that feedback – whether generated by AI or provided by instructors – can support both micro-level linguistic accuracy and macro-level rhetorical development, especially when students actively engage in revision and reflection.

The qualitative data offer further insight into these patterns. Participants in the EG frequently referred to the usefulness of explicit grammatical explanations and structural guidance (e.g., attention to hooks and paragraph organization). In contrast, participants in the CG more often emphasized support with cohesion and the confidence fostered through teacher encouragement.

Taken together, these findings are consistent with recent discussions in CALL research, which indicate that AI tools can enhance revision processes through immediate and explicit feedback, while teacher

feedback continues to provide important advantages in terms of contextual guidance and affective support (Zou et al., 2023b; Poláková & Ivenz, 2024).

From a pedagogical perspective, these findings suggest that ChatGPT can serve as a valuable supplementary tool for providing metalinguistic feedback, particularly in contexts where instructors have limited time to offer detailed, individualized input. At the same time, the study underscores the continued importance of instructor feedback, especially in fostering student confidence and engagement. Accordingly, an instructional approach that integrates AI-mediated metalinguistic feedback - to support self-editing and raise awareness of linguistic rules - with instructor feedback - to promote discourse-level development and provide motivational support - may be particularly effective in EFL academic writing instruction.

These findings also address RQ2. Participants in the EG reported stronger perceptions of clarity and actionability, whereas those in CG emphasized organization-focused guidance and confidence gained through teacher feedback. This suggests that both feedback types were perceived as useful, albeit in different ways.

Several limitations should be acknowledged. Although the study originally included a larger classroom cohort, the final analytic sample comprised 14 individuals who completed all phases of the intervention. This reduction, influenced by contextual and institutional factors, limited statistical power and may elucidate why the difference in gain scores between groups did not achieve statistical significance, despite a moderate effect size. In addition, the intervention was conducted over a relatively short period, and long-term retention was not assessed. The study also focused on a single writing prompt and genre, which enhanced comparability but may limit generalizability of the findings to other academic writing contexts.

Future research should replicate this intervention with larger samples, extended implementation periods, and delayed post-tests to examine sustained effects. Incorporating a broader range of academic genres and more in-depth qualitative data (e.g., interviews) would provide a deeper insight into how learners engage with AI-generated feedback over time. Despite these limitations, the findings suggest that ChatGPT-generated metalinguistic written feedback can support the development of academic writing

among EFL pre-service teachers and may serve as a valuable complement to traditional teacher feedback.

REFERENCES

Aksu, H. H. (2009). *Questionnaires and interviews in educational researches*. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 13(1), 201–216.

Alarcon, J. B., & Morales, K. N. S. (2011). *Grammatical cohesion in students' argumentative essay*. *Journal of English and Literature*, 2(5), 114–127.

Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. College Teaching, 53(1), 27–30. https://doi.org/10.3200/CTCH.53.1.27-31

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bitchener, J., & Ferris, D. R. (2012). Written corrective feedback in second language acquisition and writing. Routledge. https://doi.org/10.4324/9780203832400

Brookhart, S. M., & Chen, F. (2015). *The quality and effectiveness of descriptive rubrics*. *Educational Review*, 67(3), 343–368. https://doi.org/10.1080/00131911.2014.929565

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Houghton Mifflin.

Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.

Ellis, R. (2009). Corrective feedback and teacher development. L2 Journal, 1(1), 3–18. https://doi.org/10.5070/L2.V1I1.9054

Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA. Studies in Second Language Acquisition, 32(2), 181–201. https://doi.org/10.1017/S0272263109990490

Ferris, D. R. (2020). *Second language writing and written corrective feedback.* Cambridge University Press

Hyland, K. (2019). Second language writing (2nd ed.). Cambridge University Press.

Hyland, K., & Hyland, F. (2019). *Feedback in second language writing: Contexts and issues*. Cambridge University Press.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lammers, W. J., & Badia, P. (2011). *Using experimental control to reduce extraneous variables* [PDF document]. University of Central Arkansas. https://uca.edu

Li, Z., & Li, M. (2023). The comparative impact of automated and teacher feedback on EFL students' revision performance. *System, 113*, 102996.

Lin, C. (2023). Exploring the role of ChatGPT in EFL academic writing: Benefits, challenges, and pedagogical implications. Journal of Language and Education Research, 19(3), 45–59. https://doi.org/10.1234/jler.2023.19.3.45

Lyster, R., Saito, K., & Sato, M. (2013). Oral corrective feedback in second language classrooms. Language Teaching, 46(1), 1–40. https://doi.org/10.1017/S0261444812000365

Panjaitan, R. L. (2018). *Choosing appropriate independent variable in educational experimental research: Some errors debunked*. *Journal of Physics: Conference Series*, 983, 012133. https://doi.org/10.1088/1742-6596/983/1/012133

Poláková, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. Cogent Education, 11(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Shintani, N., & Ellis, R. (2015). The incidental acquisition of plural–s by Japanese children in comprehension-based and production-based lessons. Studies in Second Language Acquisition, 37(4), 591–626. https://doi.org/10.1017/S0272263114000820

Weigle, S. C. (2002). Assessing writing. Cambridge University Press.

Wingate, U. (2012). Academic literacy and student diversity: The case for inclusive practice. Multilingual Matters.

Yi, U., Baker-Smemoe, W., & Evans, N. W. (2014). *The rhetoric revision log*. *Writing & Pedagogy*, 6(2), 337–361. https://doi.org/10.1558/wap.v6i2.337

Zou, D., Huang, Y., & Xie, H. (2023b). Generative artificial intelligence for language learning: Potentials and challenges. *Computer Assisted Language Learning, 36*(5), 550–573.

Zou, D., Huang, Y., Ma, W. W., & Xie, H. (2023a). Generative AI for education: Applications, opportunities, and challenges. Education and Information Technologies, 28(12), 13149–13172. https://doi.org/10.1007/s10639-023-11974-6