

La retroalimentación generada por inteligencia artificial como  
herramienta para el desarrollo de estrategias cognitivas y  
metacognitivas en la escritura  
en inglés como lengua extranjera\*

AI-generated feedback as a tool for developing cognitive and  
metacognitive strategies in EFL writing

Mauricio Conde Torrez  
Centro de Enseñanza y Traducción de Idiomas (CETI), Universidad Mayor de San Andrés  
(UMSA),  
La Paz, Bolivia  
E-mail: [condemau182@gmail.com](mailto:condemau182@gmail.com)  
ORCID: <https://orcid.org/0000-0001-5958-4563>

Fecha de recepción: 18 de agosto de 2025  
Fecha de aceptación: 20 de octubre de 2025

\*Declaro no tener ningún tipo de conflicto de interés que haya influido en mi artículo.

**Abstract:** This study explores the use of artificial intelligence (AI) to enhance EFL students' writing by fostering cognitive and metacognitive strategy development. Conducted as a case study, it involved twenty students completing five writing tasks assessed using ChatGPT-generated corrective and formative feedback based on a TOEFL criteria rubric included within a specific prompt. A mixed-methods design was employed to measure quantitative improvements and qualitative perceptions. The results indicate improvements in writing skills, especially in Task Fulfillment and Organization in written essays, with modest progress in Grammar, Sentence Structure, and Vocabulary Use. Students reported that the feedback and the correction tasks promoted cognitive strategies such as noticing and error analysis and metacognitive strategies such as monitoring and self-evaluation, contributing to skill improvement. The study suggests that AI can support personalized instruction, assist in formative assessment, and reduce teacher workload. However, its effective implementation requires ongoing supervision to address feedback inaccuracies and ensure meaningful learning outcomes.

**Keywords:** Artificial intelligence (AI), writing skills, cognitive strategies, metacognitive strategies, AI-generated feedback.

**Resumen:** Este estudio explora el uso de la inteligencia artificial (IA) para mejorar la escritura de estudiantes de inglés, mediante el desarrollo de estrategias cognitivas y metacognitivas. Realizado como un estudio de caso, involucró a veinte estudiantes que completaron cinco tareas de escritura evaluadas mediante retroalimentación correctiva y formativa generada por ChatGPT, a partir de un *prompt* específico según criterios del examen TOEFL. Se empleó un diseño mixto para medir el progreso cuantitativamente y las percepciones cualitativamente. Los resultados indican mejoras en las habilidades de escritura, especialmente en el cumplimiento de la tarea y la organización de ensayo, con progreso modesto en gramática, estructura de oraciones y uso de vocabulario. Los estudiantes señalaron que la retroalimentación y las tareas de corrección promovieron estrategias cognitivas como el análisis de errores y la captación (*noticing*), así como estrategias metacognitivas como la supervisión y la autoevaluación. El estudio sugiere que la IA puede apoyar la instrucción personalizada, facilitar la evaluación formativa y reducir la carga laboral docente. No obstante, su implementación efectiva requiere supervisión continua para abordar posibles imprecisiones en la retroalimentación y garantizar resultados de aprendizaje significativos.

**Palabras clave:** inteligencia artificial (IA), habilidades de escritura, estrategias cognitivas, estrategias metacognitivas, retroalimentación generada por IA.

## INTRODUCTION

Mastering writing skills is essential to achieving higher proficiency levels, as most higher education institutions globally require the submission of various types of written works in English (Adriansen et al., 2023). Writing paragraphs, essays, articles, and other kinds of texts requires not only linguistic knowledge but also careful thought about organizing ideas, drafting, revising, and rewriting (Zemach & Rumisek, 2005). However, many students fail to produce adequate texts due to a lack of awareness of these elements. One major limitation is that teachers do not have enough time to review, check, and provide formative feedback on students' writing, especially in large classes (Kaufer & Ishizaki, 2025). Alternatively, the sheer volume of course content often leaves insufficient time to reinforce and refine students' writing skills. **Consequently, students fail to fully develop these skills and struggle to guide their own learning or compensate for skill deficiencies.** As a result, **educators** must find new methods and techniques to improve and personalize learning. Recently, artificial intelligence (AI) has emerged as a promising tool to **address this challenge.**

AI is transforming various fields, including language teaching, which is evident in the increasing research interest from 2019 to the present date. In a systematic review of 95 articles related to AI and education by Tan et al. (2025), English language teaching was the first research topic, followed by computer science and art. This illustrates the growing trend of developing new teaching strategies that implement AI. This technology offers teachers new benefits, such as automation in grading tests and giving feedback (Cohen et al., 2024). Also, it can be used to adapt learning materials and personalize education (He, 2024). This suggests that these technologies can encourage students to develop autonomy if used properly.

The adoption of AI also presents implementation challenges, notably the increasing student use of AI for writing, with or without educators' consent. In a study conducted with 83 high school students, Díaz-Arce (2023) found that two-thirds of the students had used AI for essays, primarily for paraphrasing text or generating initial ideas. This trend raises concerns about academic integrity, plagiarism, and the potential erosion of critical thinking skills due to over-reliance on the technology. Compounding this, most educators are unfamiliar with the technology; for instance, a University of Miami study (Cohen et al., 2024) found that 60% of faculty had never

used AI. In the Bolivian context, more research on this topic is needed, since AI might significantly improve the quality of education with appropriate implementation. Otherwise, a failure to act could widen the educational inequality gap, especially given that most educational systems worldwide are projecting its use at all levels. Consequently, international institutions are recommending its implementation in education, emphasizing that teachers should promote its adequate and ethical use (UNESCO, 2022).

This study explores the role of ChatGPT-generated corrective and formative feedback to improve writing proficiency by fostering cognitive and metacognitive strategies. A seven-week longitudinal case study was conducted, tracking 20 students' performance across five writing tasks. A mixed-methods approach was employed to analyze the efficacy of the feedback and students' perceptions of it. The findings aim to contribute to the understanding of the role of AI in language learning.

## LITERATURE REVIEW

### **1. Writing and the Development of Cognitive and Metacognitive Strategies**

Most English language programs and international tests require students to write opinion essays at advanced levels. However, writing an essay at advanced proficiency levels goes beyond grammatical accuracy. Writing involves managing multiple cognitive processes, such as planning to express ideas in texts, refining sentences, and reviewing texts (Kaufer & Ishizaki, 2025). "Therefore, students need to be aware of these processes to enhance their performance. To do this, instructors must consider the cognitive aspects of writing skills to promote learner autonomy, which means taking responsibility for one's learning. In this regard, behavioral autonomy involves the use of cognitive and metacognitive strategies to take control of the learning process (Benson, 1997). The focus on cognitive processing in language learning has gained importance, being recognized as a key element within the Common European Framework of Reference for Languages (CEFR) guidelines for instructional design and lesson planning (Nagai *et al.*, 2020).

Learning is a conscious thinking process involving the deliberate use of learning strategies, an approach known as strategy-based instruction. The main learning strategies can be cognitive, metacognitive, and social-affective (Lakshmi Priya & Priyadarshini, 2021). Cognitive strategies are oriented to tasks in which attention plays a crucial role, since nothing can be learned or

even improved if it has not been noticed before (Lightbown & Spada, 2006). Attention can be directed through language-focused tasks to enhance deep processing (Nation, 2007).

In writing tasks, students need to answer the given question or prompt with a good flow of ideas, a clear organization, and the use of complex grammar structures and vocabulary. The main aspects assessed in international tests, such as the TOEFL, are sentence structure, cohesion, coherence, the use of appropriate vocabulary, and the ability to communicate a well-supported opinion (Phillips, 2006). Often, students focus on only one aspect and neglect the others. **Therefore, effective writing instruction must provide frequent practice to direct students' attention to all these cognitive aspects.** This approach enables students to gain an overall perspective and **reduces the cognitive load necessary to significantly improve writing skills** (Kaufer & Ishizaki, 2025). With enough practice, students might internalize this knowledge and turn it into a procedural skill, an essential aspect of communicative competence and fluency (Schmidt, 1992). It means that they could write better texts automatically, focusing on broader aspects such as expressing their ideas, rather than just on grammar, vocabulary, or style.

Learner autonomy also demands self-reflection and self-assessment of progress. Metacognition refers to the deliberate reflection and planning of learning to monitor and assess progress. To encourage metacognition, the instructor might explicitly teach writing strategies and show assessment guidelines so students know what is expected of them (Lakshmi Priya & Priyadarshini, 2021). To promote metacognition in the class, Winslow & Shaw (2017) made students read articles about metacognition and discuss them in front of the class to increase awareness of its importance in a writing course at the Rochester Institute of Technology. In explicit instruction, the instructor might show the rubric used to assess writing tasks, such as those in the Common European Framework of Reference for Languages (CEFR) or TOEFL tests. Aligning the tasks to standard descriptors or rubrics helps to guide the improvement of written skills (Nagai *et al.*, 2020). In advanced courses, students should know the different competencies for B2, C1, or C2 that they need to develop so they can direct their learning. Instructors must include demanding tasks according to the expected level. Technology, such as artificial intelligence, might assist in designing rubrics and tasks aligned with these standards (He, 2024). Another way to promote metacognition is to

provide formative feedback on students' written work and direct their attention to areas that need improvement. In this regard, AI can also be an important additional source of detailed feedback, which could encourage attention and reflection on the aspects they need to improve.

## **2. AI and Writing Practice**

Artificial intelligence (AI) can assist students in improving writing production, since it is based on language analysis. This technology uses Large Language Models (LLMs), which analyze the frequency and distribution of a large quantity of words, sentences, and phrases (Ji, 2024). This advancement was developed thanks to the development of machine learning, which enabled the training of AI with vast amounts of language use instances from literature, magazines, web pages, and more (Luchs, 2023). Therefore, generative AI can generate novel grammatical constructions and text with appropriate discourse for any context or topic. These features promise many benefits for education, especially language learning.

New technologies have always revolutionized the English language teaching field. This was the case with the implementation of radios to enhance listening comprehension and the extended multimodal learning that mobile phones offer, resulting in the development of the Mobile Assisted-Language Learning approach (Pyo & Lee, 2024). AI promises to reduce the teachers' workload and personalize education, as it can be automated to give feedback and grade quizzes and texts. Materials might be created automatically, and teacher planning could be fostered since this technology can suggest language learning activities (Cohen *et al.*, 2024). AI could also personalize education and enhance the development of language skills, acting as a tutor. Waly & Zakiyyah (2024) studied its implementation to foster communicative competence in an ESP course in Indonesia. In the survey, 77% of participants showed a positive attitude towards AI tools, highlighting their benefits in giving feedback for oral communication and gaining insights for projects, papers, or other written tasks. However, 64% of participants were also concerned about the impact on critical thinking skills if used constantly.

The impact of AI tools in writing has been studied under the emerging Automated Writing Corrective Feedback (AWCF) approach by using AI-powered programs and even chatbots. Sanosi (2022) analyzed the impact of AWCF by using Grammarly with 64 students from the English Language

Literature department at Prince Sattam bin Abdulaziz University in Saudi Arabia. Grammarly is an AI-powered writing assistant that can give suggestions and feedback about word choice, grammar, and sentence structure. The students who used Grammarly showed improvement mainly in article usage, subject-verb agreement, and the use of singular and plural forms. The use of verb tenses was difficult to improve since it is related to semantic aspects. A similar study using another AI-powered writing assistant was carried out by Al-Gaithi & Behforouz (2025). They used EditGPT, which is a plugin from Google that highlights and shows suggestions and inconsistencies to improve writing. In this experimental study, 60 students were divided into three groups, with only one experimental group that received detailed feedback and used the AI-powered tool. The students who used this tool performed better than the other groups.

Other studies focus on the use of chatbots such as ChatGPT, an AI-powered chatbot that leverages Large Language Models (LLMs) and machine learning to generate, understand, and analyze text. Boudouaia et al. (2024) studied the effects of using ChatGPT for writing in students from a private EFL program in Algeria. The students were divided into a control and an experimental group, with 76 students in total. The experimental group received feedback from ChatGPT-4. The results indicated they developed a more effective sense of language and demonstrated more consistent use of grammar and vocabulary. A similar study by Pipia & Gurgenshvili (2025) was carried out at Georgian universities, involving six groups totaling 33 upper-intermediate students. The experimental group was instructed about the use of AI and integrated the feedback from ChatGPT. Both groups received feedback from the teacher. Students who used AI tools showed improvement in their organizational skills, coherence, grammar, vocabulary, and spelling. However, overdependence was a main concern. While the results showed that the pretest outperformed the posttest of these groups, there were nevertheless improvements, especially in meeting the demands of the writing task and higher levels of writing proficiency.

Chatbots can be used to enhance writing skills due to their accessibility, since they can be used from mobile devices anywhere. They can also act as tutors by offering suggestions and providing formative feedback on students' performance. Most students already use them for grammar revision and explanations of complex concepts, especially in ESP courses (Savitri *et al.*, 2025). That is why their use should be modeled and guided by the teacher. There are concerns about the use of generative AI in writing, as students may create

texts containing false information. However, there are no ethical concerns about plagiarism or fake content if students use it to edit and get feedback on their writing. Overdependence should be reduced and controlled by the teacher.

## METHODOLOGY

### 1. Research Design

This study employed a longitudinal case study design, which is useful to gain a more in-depth understanding of the progression and the specific properties of the studied phenomenon (Hernandez *et al.*, 2014). In order to assess the impact of AI implementation and feedback, the performance of a single cohort was tracked over seven weeks, examining trends of improvement or stagnation in writing performance across five time points. In addition, to gain deeper insights into the subjective dimensions of writing skills development, this study employed a mixed-methods approach, specifically incorporating a sequential explanatory design. In this design, quantitative data are gathered first, followed by qualitative data to support and analyze the findings in depth (Vizcaíno *et al.*, 2023). In the quantitative phase, students' essays were graded by ChatGPT-4, which generated automated scores that were recorded in a progress sheet to monitor performance over time. In the qualitative phase, students completed a reflective questionnaire that explored their experiences with AI-generated formative feedback and their perceptions of progress.

### 2. Participants

Twenty EFL upper-intermediate (B2) learners enrolled in an intensive course participated in this study. The study was conducted at Centro de Enseñanza y Traducción de Idiomas, which is part of the Department of Linguistics and Languages at Universidad Mayor de San Andrés. The course encompassed five weekly sessions (three hours per day) over seven weeks during the first semester of 2025. Convenience sampling was used to select the participants, as they comprised the researcher's assigned class. All the participants met the study's inclusion criteria by providing informed consent and completing all five required essays.

### 3. Instruments

**Progress Sheet:** Five writing tasks were completed in the class during the final five weeks of the course. The writing tasks required students to write



an essay in class without access to devices, dictionaries, or external sources. After finishing the essay, students had to type and submit it through a curated prompt provided by the researcher. This prompt ensured the generation of specific Automated Written Corrective Feedback (AWCF) and formative feedback from ChatGPT, which graded the essays using a fixed rubric based on TOEFL writing test criteria. The same prompt was used across all the writing tasks. The full prompt appears in Appendix A.

**Scores** of the essays were recorded in a progress sheet in order to quantify performance across the following categories:

- Task Fulfillment
- Organization and Development
- Grammar and Sentence Structure
- Vocabulary Use
- Cohesion and Coherence

**The progress sheet** served to track students' progression across the five writing tasks and to identify performance trends. Additionally, the progress sheets fostered cognitive and metacognitive awareness by including a section where students would write the main aspects they thought they needed to improve, based on the feedback received from each writing task.

**Reflection questionnaire:** A questionnaire with three open-ended questions was delivered via Google Forms at the end of the course in order to gather students' perceptions of using ChatGPT for writing improvement. The instrument aimed to capture how students experienced the AI feedback process and whether they believed it contributed to the development of writing skills through cognitive and metacognitive strategies, as they were asked to reflect on the effectiveness of AI in their learning. This reflective process enabled a deeper analysis of the cognitive and metacognitive processes involved in writing skills improvement, which are central to this study.

#### 4. Procedures

Students wrote five argumentative essays during the final five weeks of the course. Each writing task followed a TOEFL-like format, requiring a 250-word written response to a question on paper, completed within twenty-five minutes during class without external assistance. According to CEFR descriptors, B2-level students are expected to write essays on a wide range of

topics, giving reasons for or against a particular point of view (Nagai *et al.*, 2020). Consequently, the topics were related to creativity, resilience, ethics, career choices, and leisure time. The essays were handwritten and submitted to the researcher for registration, ensuring that no modifications were made after the task. Then, the students transcribed their original essays into ChatGPT using the standardized prompt provided. The chatbot generated three outputs: (1) scores and formative feedback for each rubric category; (2) a list of major grammar errors with explanations; and (3) a full corrected version of the essay.

Students were instructed to handwrite all the feedback and corrections into their language portfolios. This process aimed to foster cognitive awareness by drawing attention to strengths and weaknesses, so students would try to improve in the next writing tasks. In addition, it encouraged metacognitive awareness, as students had to read, analyze, and reflect on their performance. Progress was tracked through the scores recorded in the progress sheet, which served to gather quantitative data and identify trends in progression or stagnation. The use of a consistent prompt ensured that the feedback remained aligned with specific guidelines, reducing variations in the feedback provided by the chatbot.

## RESULTS

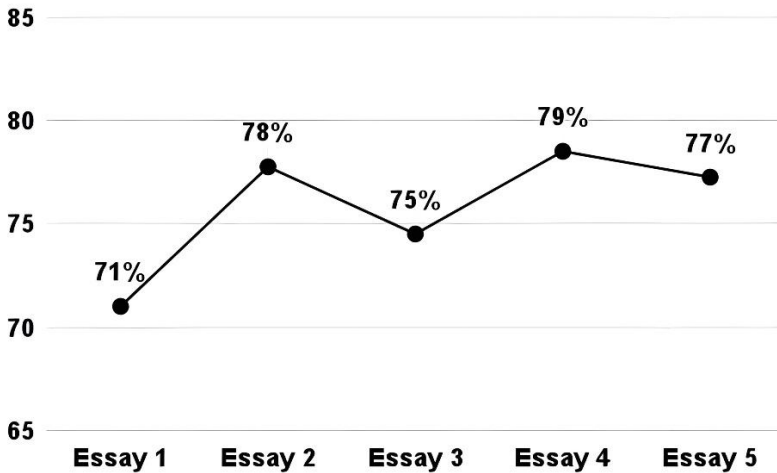
### 1. Writing Tasks

The quantitative data were obtained from the feedback and assessment provided by ChatGPT after students submitted each of their five essays. The assessment was based on five criteria: task fulfillment, organization and development, grammar and sentence structure, vocabulary use, and cohesion and coherence. Each category contributed 20% to the final score. Each evaluation was recorded on an individual progress sheet to track and compare the students' performance across the five writing tasks. Then, the average scores per category and the overall score were calculated and analyzed using Google Sheets. To identify performance trends, the results are presented in figures. Each figure corresponds to one of the five writing categories, followed by a summary figure that shows the overall writing performance across the five essays.

#### 1.1. Task Fulfillment

The progression of students' performance in the Task Fulfillment category across the five writing tasks is shown in Figure 1.

Figure 1: *Progress in Task Fulfillment across Writing Tasks*



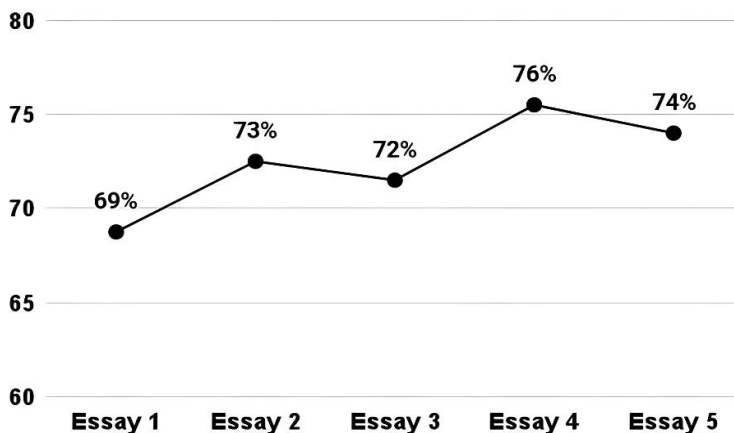
Source: The author's elaboration is based on the data gathered.

The first essay had an average score of 71%, suggesting that students had difficulty addressing the writing question, possibly due to unfamiliarity with the question format. In Essay 2, performance improved to 78%. Scores decreased slightly in Essay 3 (75%), but increased again in Essay 4, which was the highest average (79%). Essay 5 shows a slight decline (77%), but overall, the trend suggests a gradual improvement in students' ability to fulfill the task requirements.

### *1.2. Organization and Development*

Figure 2 illustrates the progression in students' scores for organization and development across the five writing tasks.

Figure 2: *Progress in Organization and Development across Writing Tasks*



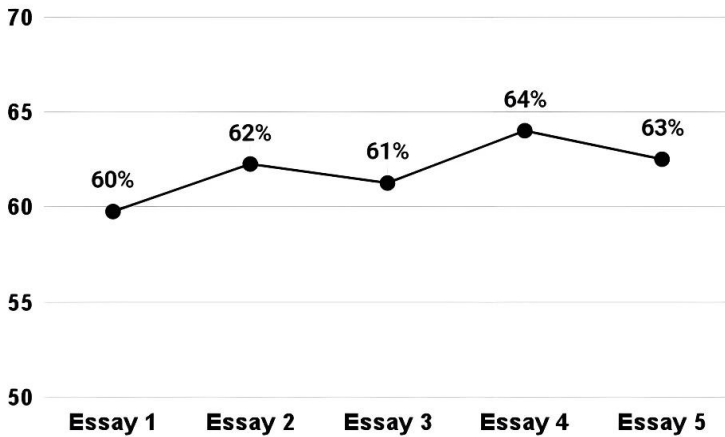
Source: The author's elaboration is based on the data gathered.

The average score for Essay 1 was 69%, and it increased to 73% in Essay 2. There was a slight decline in Essay 3 (72%), but scores increased in Essay 4, reaching 76%. In the final task, there was a small drop to 74%. Despite the changing percentages, there is a tendency towards improvement over the five tasks.

### *1.3. Grammar and Sentence Structure*

The Grammar and Sentence Structure category was one of the weakest in students' overall performance. As shown in Figure 3, the average scores did not reach 70%, unlike the other categories. There was a trend of progression, starting with 60% in Essay 1, 62% in Essay 2, and a slight decrease to 61% in Essay 3. Essay 4 reached its peak with 64%, followed by a slight decrease to 63% in Essay 5. In general, the figure shows modest improvement in this category.

Figure 3: *Progress in Grammar and Sentence Structure across Writing Tasks*

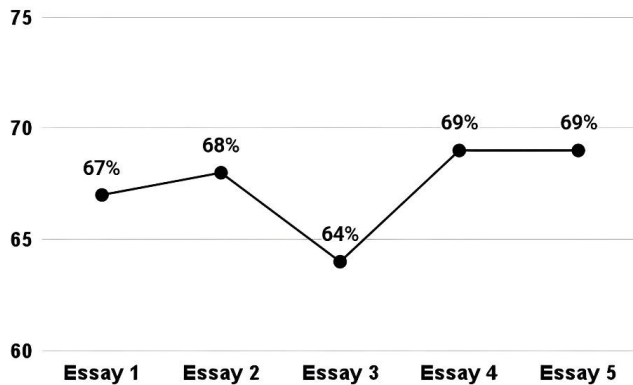


Source: The author's elaboration is based on the data gathered.

#### 1.4. Vocabulary Use

There is a slight tendency towards progression as shown in Figure 4.

Figure 4: *Progress in Vocabulary Use across Writing Tasks*



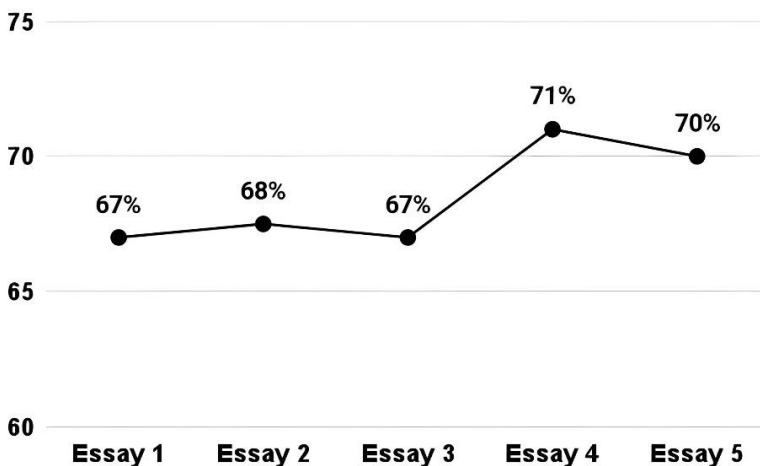
Source: The author's elaboration is based on the data gathered.

The average score was 67% in Essay 1. In Essay 2, it increased to 68%, but decreased to 64% in Essay 3, which was the lowest score. In Essays 4 and 5 remained stable at 69%. The decrease in Essay 3 might be attributed to the complexity of the topic, which addressed ethical and moral issues.

### 1.5. Cohesion and Coherence

Figure 5 shows the starting average score of 67% in Essay 1, with a slight increase in Essay 2 to 68%. It went back to 67% in Essay 3. Essay 4 was the peak with 71%, followed by a decrease to 70% in Essay 5. There is a slight tendency to improve cohesion and coherence, but the trend is modest.

Figure 5: *Progress in Cohesion and Coherence across Writing Tasks*



Source: The author's elaboration is based on the data gathered.

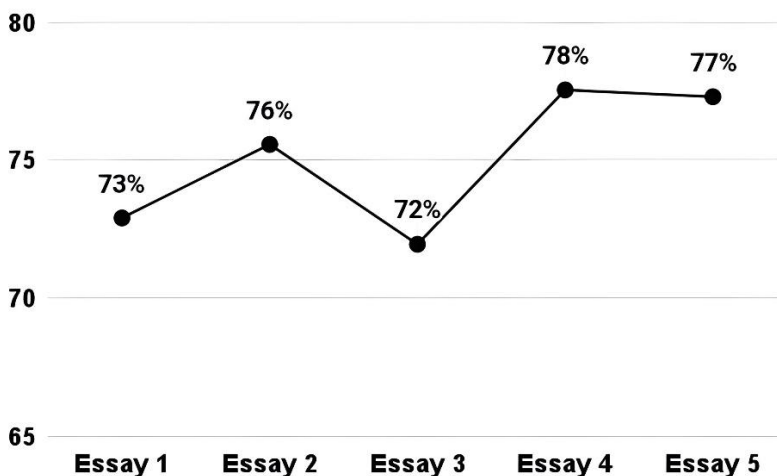
### 1.6. Overall writing progress

Figure 6 illustrates the total scores obtained in each writing task throughout the course, with each score presented as a percentage. The figure shows that there is an overall trend towards progression in students' essays.

The entry level for Essay 1 was 73%, followed by an increase to 76% in Essay 2. Essay 3 was the lowest across all the essays, with 72%.

This decline might be linked to the demanding topic of this essay. Essay 4 shows an increase to 78%, and finally, Essay 5 scored 77%, which shows a slight decrease.

Figure 6: *Total Essay Scores Across Writing Tasks*



Source: The author's elaboration is based on the data gathered.

### *1.7. Reflection questionnaire*

After completing the five essays, students received a questionnaire with three open-ended questions via Google Forms to gather information about their perceptions of using ChatGPT for writing improvement. A thematic analysis of the responses was conducted, and the findings were coded using Atlas.ti. The identified categories are shown according to their frequency of appearance.

#### ***Question 1. How did the use of AI let you think about your own mistakes in writing?***

Three main categories were identified based on students' responses, as summarized in Table 1.

Table 1: Identified Categories from the First Question in the Questionnaire

Category	Definition	Frequency	Percentage
Noticing mistakes	Students directed their attention to their own mistakes in grammar, vocabulary, organization, etc.	15	75%
AI as tutor	AI gave examples, suggestions, and explanations to guide student improvement	7	35%
Corrective feedback	AI provided corrections and reformulations to linguistic errors	3	15%

Source: The author's elaboration is based on the data gathered.

Most students (75%) reported that using ChatGPT encouraged them to notice errors in grammar, vocabulary, and organization, among others. Seven students (35%) also highlighted the value of AI as a tutor, since it offered suggestions, examples, and explanations that supported them to improve. Finally, three students specifically emphasized that the corrective feedback had been useful to improve their errors.

### ***Question 2. Did you see progress in your writing performance?***

This question was metacognitive, intended to encourage students to reflect on their own performance and learning process (self-evaluation). The students' responses are summarized in Table 2.

Table 2: Students' Self-evaluation of their Progress

Category	Definition	Frequency	Percentage
Yes (perceived progress)	Student reports progress (e.g., more vocabulary, better structure)	15	70%
No (No progress)	Student sees no change or feels stuck	3	15%
Limited progress	Partial or inconsistent change (e.g., "same scores")	3	15%

Source: The author's elaboration is based on the data gathered.



Fifteen students (70%) reported that they felt they had progressed using AI, especially because they were able to notice their errors, expand their vocabulary, improve the use of grammar structures, and gain more confidence. Three students (15%) stated that they did not feel they had progressed, as they did not observe improvement in their essay scores. One student reported feeling stressed, and another mentioned the importance of improving the current language proficiency. Finally, the other three students (15%) mentioned that they had seen little progress because their errors in grammar or vocabulary persisted across all essays.

**Question 3.** *After using AI, what are some disadvantages of using AI to improve writing?*

This question analyzed perceptions about the limitations of using AI. The findings are summarized in Table 3. Six students (40%) highlighted that becoming dependent on using AI was their main concern, especially when used excessively. Another six students (40%) reported different challenges when interacting with these technologies. Some students mentioned that they needed more language proficiency to interact with AI, while others expressed feeling frustrated or anxious, probably due to the continuous corrections. Others mentioned feeling unnatural and impersonal in their interaction with these chatbots. Four students (20%) mentioned that they did not see disadvantages when using this technology. Finally, the other four students (20%) stated that AI was not always reliable, as it had limitations in understanding the context or changing ideas when offering suggestions.

Table 3: *Perceived Disadvantages of Using AI*

Category	Definition	Frequency	Percentage
Overdependence	Concerns about becoming too dependent on AI	6	40%
Interaction barriers	Problems when interacting with AI	6	40%
No disadvantage	No negative aspects perceived	4	20%
AI unreliability	doubts about AI's accuracy or helpfulness	4	20%

Source: The author's elaboration is based on the data gathered.

## DISCUSSION

This study explored the effectiveness of AI-generated feedback in developing students' cognitive and metacognitive strategies and promoting progress in essay writing. The findings suggest that the participants improved their essay writing skills across the five different writing tasks. As illustrated in Figure 6, which tracks the total essay scores across tasks, progress followed a U-shaped curve. This pattern is consistent with the discussion on Schmidt's (1992) restructuring hypothesis of accuracy and fluency, which posits a temporary decline in performance when students face tasks that require restructuring their knowledge. Once new procedures are internalized, there is an increase in their performance. Before the study, most students wrote essays without specific attention to criteria like task fulfillment, grammar, and sentence structure. The continuous feedback likely prompted participants to restructure their cognitive approach to better address the complex demands of the writing tasks.

The writing assignments were part of a structured process designed to encourage students to engage more actively with the feedback. The development of cognitive and metacognitive strategies is a cyclical process, since a person needs to notice and reflect on their performance to develop new strategies for improvement. Attention plays an important role in developing cognitive skills (Lakshmi Priya & Priyadarshini, 2021). In this study, directing attention through AI-generated corrective and formative feedback towards specific errors and weaknesses may have fostered a more holistic awareness that supported error analysis and correction. To promote noticing, analysis, and correction of errors, students were asked to handwrite and present all the formative feedback received from the different rubric categories, compile a list of grammatical errors with the corrections, and submit the fully corrected essay.

Reading the feedback briefly does not guarantee that students will deeply process or understand it, though. To promote active engagement, higher-order thinking skills must be encouraged, as students need to understand the information, analyze it, and generate output that integrates all they have read (Üstünlüoğlu, 2004). The AI-generated formative feedback encouraged self-reflection and the refinement of metacognitive strategies, which led to progression. Therefore, metacognitive awareness was promoted by making students write the scores on their progress sheet and identify and summarize the highlights from the formative feedback obtained, as it prompted self-

evaluation by reading, processing, interpreting, and synthesizing the information about their performance. This cyclical process contributed to developing and improving the strategies that they would use for the next writing task.

The majority of students reported through the questionnaire that using the AI tutor increased their awareness of their own writing flaws and mistakes, a level of self-reflection they had not previously practiced. The reported anxiety some students experienced when receiving performance scores or noting marginal improvement across tasks suggests a developing metacognitive awareness. Despite the persistent flaws in grammar, vocabulary, and other areas, the majority of students still perceived that they had progressed in their writing skills.

The main areas where there was significant improvement were task fulfillment and cohesion and coherence in essays. This aligns with the results of Pipia & Gurgenshvili (2025), in which students also improved their organizational skills and the coherence of their writings. Also, these findings align with Boudouaia *et al.* (2024), who observed that students improved their ability to meet the demands of writing tasks. Grammar and sentence structure, as well as vocabulary use, did not show significant progress, similar to the research carried out by Sanosi (2022). According to this author, there are elements in grammar, such as verb tenses, that are not only related to the form but also to the meaning and context in which they are used. Students probably lacked enough grammatical accuracy knowledge before the course. Therefore, this is an issue that might take longer for students to restructure and correct. However, students did value AWCF, stating that it made them notice their grammar and writing mistakes, which is the first step towards restructuring and improving their accuracy (Lightbown & Spada, 2006).

The progress and positive views on AI's corrective and formative feedback suggest it can be a valuable tool for personalizing education and promoting learner autonomy (Cohen et al., 2024). However, educators implementing this technology must address key issues. One primary concern, central to most research on AI in education, is student overdependence (Pipia & Gurgenshvili, 2025; Savitri et al., 2025; Waly & Zakiyyah, 2024).

In the context of this study, this risk was mitigated: students used AI only for correction, editing, and feedback, not for generating essays, and were required to write their essays in class, ensuring they remained the main agents in the writing process. Furthermore, students were explicitly instructed that

these tools assist in identifying weaknesses, but that they, as writers, were ultimately responsible for deciding what to keep and revise. This guidance was successful, as reflected in questionnaire responses where students noted that AI is not always reliable, occasionally failing to understand context fully or altering their original ideas. These critical comments indicate that students were actively developing critical thinking toward the generated information and the appropriate use of this technology.

## CONCLUSION

This study explored the use of AI-generated corrective and formative feedback to foster cognitive and metacognitive strategies in writing. The findings of this study suggest moderate progress in students' writing performance, especially in task fulfillment, organization, cohesion, and coherence, which indicate the development of cognitive strategies to improve performance. On the other hand, grammar, sentence structure, and vocabulary remained stable, possibly because these aspects require more time to be cognitively restructured before observing notable improvements. However, the questionnaire responses revealed that the feedback from ChatGPT had encouraged students to notice their grammatical errors and vocabulary use, fostering metacognitive awareness.

These findings correlate with several current pedagogical implications regarding the use of AI in education. First, AI can be used to personalize instruction when used as a tutor to promote learner autonomy and reflection on performance. It can also be used to correct and grade writing tasks in many aspects, such as grammar, sentence structure, vocabulary use, organization, and more. This automated correction might reduce the teacher's load, especially in correcting writing tasks in classes with a lot of students. However, this technology is still limited and requires teacher supervision. An effective way is to curate and give specific prompts based on specific learning objectives and instructional content. In this way, each student might receive personalized corrective and formative feedback while engaging in the same learning tasks.

This study had some limitations that should be considered for future research. The course lasted seven weeks, and students wrote five essays. Future research might focus on longer periods, such as a full semester, with more writing tasks to be tracked and analyzed. With more time and writing samples, a deeper understanding of AI effectiveness in developing writing

skills can be obtained. The number of participants was another limitation of this study, so a case study was conducted. It might be beneficial to conduct longitudinal experimental research in the Bolivian context, incorporating both an experimental and a control group in order to gain deeper insights into students' progress and the factors that may foster it. The reliability of AI is another factor to consider, as this technology is still being developed and may make errors. Allowing AI to score students' assignments should be carefully monitored by the teacher. Students might report inaccuracies in the examples or grading, so the instructor must address these concerns and be in charge of the grading. Therefore, future research might focus on evaluating the effectiveness of AI-generated feedback and scoring of written assignments.

This technology is transforming all aspects of society and the field of education, so more research and critical reflection on its appropriate implementation are needed, especially in the Bolivian context, to empower students to use it to guide their learning. In this regard, this study aims to encourage further research on the development of writing skills by implementing AI in the classroom.

## REFERENCES

- Adriansen, H. K., Juul-Wiese, T., Madsen, L. M., Saarinen, T., Spangler, V., & Waters, J. L. (2023). Emplacing English as lingua franca in international higher education: A spatial perspective on linguistic diversity. *Population, Space and Place*, 29(14), 1–14. <https://doi.org/10.1002/psp.2619>
- Al-gaithi, A., & Behforouz, B. (2025). Effects of interaction with AI-assisted writing evaluation }on EFL students' writing performance. *Knowledge Management & E-Learning*, 17(2), 206–224. <https://doi.org/10.34105/j.kmel.2025.17.009>
- Benson, P. (1997). The philosophy and politics of learner autonomy. En P. Benson & P. Voller (Eds.), *Autonomy and Independence in Language Learning*. Longman.
- Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on Chatgpt 4 as an innovative approach to enhancing English as a Foreign Language writing. *Journal of Educational Computing Research*, 62(6), 1–29. <https://doi.org/10.1177/07356331241247465>
- Cohen, S., Mompelat, L., Mann, A., & Connors, L. (2024). The linguistic

- leap: Understanding, evaluating, and integrating AI in language education. *Journal of Language Teaching*, 4(2), 23–31. <https://doi.org/10.54475/jlt.2024.012>
- Díaz-Arce, D. (2023). Plagio a la Inteligencia Artificial en estudiantes de bachillerato: un problema real. *Revista Innova Educación*, 5(2), 108–116.
- He, X. (2024). Enhancing Reading Comprehension with AI-Generated Adaptive Texts. *International Journal of New Developments in Education*, 6(7), 46–52. <https://doi.org/doi:10.25236/IJNDE.2024.060708>.
- Hernandez, R., Fernandez Collado, C., & Baptista Lucio, M. del P. (2014). *Metodología de la investigación* (6º). McGraw Hill.
- Ji, E. Y. (2024). Large Language Models: A Historical and Sociocultural Perspective. *Progress & Puzzles of Cognitive Science*, 48(3), e13430. <https://doi.org/10.1111/cogs.13430>
- Kaufer, D., & Ishizaki, S. (2025). *Future of writing & writing instruction in the disciplines and professions*. Carnegie Mellon University.
- Lakshmi Priya, K., & Priyadarshini, M. S. (2021). Effective Cognitive Strategies in English Language Learning. *International Journal of Emerging Technologies and Innovative Research*, 8(12), e732–e737.
- Lightbown, P., & Spada, N. M. (2006). *How languages are learned*. Oxford University Press.
- Luchs, I. (2023). AI for All? Challenging the Democratization of Machine Learning. *A peer-reviewed journal about*, 12(1), 135–147. <https://doi.org/10.7146/aprja.v12i1.140445>
- Nagai, N., Birch, G. C., Bower, J. V., & Schmidt, M. G. (2020). *CEFR-informed Learning, Teaching and Assessment: A Practical Guide*. Springer.
- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1(1), 2–12.
- Phillips, D. (2006). *Longman preparation course for the TOEFL iBT test*. Pearson Education.
- Pipia, E., & Gurgenshvili, T. (2025). The impact of Chatgpt on English as a Foreign Language learners' writing skills. An experimental study at Georgian Universities. *European Scientific Journal, ESJ*, 39(429). <https://eujournal.org/index.php/esj/article/view/19269>
- Pyo, J., & Lee, C. H. (2024). Developing learner autonomy and FL listening skills through mobile-assisted blended learning. *Innovation in*

- Language Learning and Teaching*. <https://doi.org/10.1080/17501229.2024.2372068>
- Sanosi, A. B. (2022). The impact of automated written corrective feedback on EFL learners' academic writing accuracy. *Journal of Teaching English for Specific and Academic Purposes*, 10(2), 301–317.
- Savitri, W. E., Nugroho, H. A., Munir, A., Pusparini, R., Kurniasih, E., & Arinda, D. (2025). Perspectives of pre-service English teachers on AI chatbots in ESP worksheet development. *JOLLT Journal of Languages and Language Teaching*, 13(1), 425–435. <https://doi.org/10.33394/jollt.v13i1.12964>
- Schmidt, R. (1992). Psychological Mechanisms Underlying Second Language Fluency. *Studies in Second Language Acquisition*, 14, 357–385.
- Tan, X., Cheng, G., & Ho Ling, M. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*, 8. <https://doi.org/10.1016/j.caeai.2024.100355>
- UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial*. United Nations Educational, Scientific and Cultural Organization. [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa)
- Üstünlüoğlu, E. (2004). Language teaching through critical thinking and self-awareness. *English Teaching Forum*, 3(42), 2–7.
- Vizcaino, P. I., Cedeño, R., & Maldonado, I. (2023). Metodología de la investigación científica: guía práctica. *Ciencia Latina Revista Científica Multidisciplinar*, 7(4), 9723–9762. [https://doi.org/10.37811/cl\\_rcm.v7i4.7658](https://doi.org/10.37811/cl_rcm.v7i4.7658)
- Waly, M. M., & Zakiyyah, A. M. (2024). Artificial intelligence assistance in English for communication competence of English for specific purposes. *INSANIA : Jurnal Pemikiran Alternatif Kependidikan*, 29(2), 236–256. <https://doi.org/10.24090/insania.v29i2.11140>
- Winslow, D., & Shaw, P. (2017). Teaching metacognition to reinforce agency and transfer in course-linked first-year courses. En P. Portanova, J. M. Rifenburg, & D. Roen (Eds.), *Contemporary perspectives on cognition and writing* (pp. 191–209). The WAC Clearinghouse. <https://doi.org/10.37514/PER-B.2017.0032>
- Zemach, D. E., & Rumisek, L. A. (2005). *Academic writing: From paragraph to essay*. Macmillan.



## Appendix A: ChatGPT Prompt

### *Corrective and formative feedback*

Grade my essay out of 100 using this B2 rubric based on TOEFL criteria.

Consider grammar, organization, cohesion, and whether my response meets the expected length for a B2-level essay (200-250 words).

Question:

*“Some people think that parents should plan their children’s leisure time carefully. Other people believe? Those children should decide for themselves how to spend their free time. Which idea do you agree with? Give reasons for your choice.”*

### **B2 Writing Rubric (100 points total)**

**1. Task Fulfillment** (20 points) – Does my essay fully answer the question and meet the expected length for B2?

18-20: Fully answers the question with clear ideas and examples.

Appropriate length for B2 (200-250 words).

15-17: Mostly answers the question but lacks some details. Slightly too short or too long for B2.

10-14: Partially developed; noticeably too short or too long for B2. Some ideas are missing or too wordy.

1-9: Severely underdeveloped or excessively long.

**2. Organization & Development** (20 points) – Is my essay well-structured and developed?

18-20: Clear introduction, body paragraphs with strong supporting details, and a well-rounded conclusion.

15-17: Mostly well-structured, but some ideas need more development.

10-14: Some disorganized or underdeveloped ideas.

1-9: Lacks organization; difficult to follow.

**3. Grammar & Sentence Structure** (20 points) – Are my sentences grammatically correct and well-formed?

18-20: Few or no grammar errors; varied sentence structures.

15-17: Some minor errors, but meaning is clear.

10-14: Frequent errors that sometimes affect meaning.

1-9: Many errors that make understanding difficult.

**4. Vocabulary Use** (20 points) – Is my vocabulary appropriate and varied?

18-20: Precise, natural, and topic-appropriate word choice.

15-17: Mostly appropriate, with minor awkwardness.



10-14: Limited vocabulary; some words used incorrectly.

1-9: Poor vocabulary, with many unclear words.

**5. Cohesion & Coherence** (20 points) – Are my ideas well-connected and easy to follow?

18-20: Smooth transitions, good use of linking words, logical flow.

15-17: Mostly clear, but some transitions are weak.

10-14: Some disjointed ideas; weak or missing transitions.

1-9: Hard to follow, with little connection between ideas.

Total: \_\_/100

**Correction Task (Handwritten)**

1. List my grammar, vocabulary, or punctuation mistakes one by one.

2. Write the corrected version of each mistake.

3. Provide the full corrected essay at the end.

Here's my essay: