



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 8, 103995 (2026)

<https://doi.org/10.29140/tltl.2026.103995>

Exploring the Effects of AI-Generated Feedback on EFL Writing Fluency and Quality: A Process and Product-Based Perspective



OUMAIMA BELGAMRA 

Faculty of Educational Sciences, Mohammed V University, Rabat, Morocco
belgamraoumaima0@gmail.com

Abstract

As artificial intelligence (AI) tools become increasingly embedded in educational contexts, their role in supporting EFL writing development warrants closer empirical examination. Grounded in Flower and Hayes' Cognitive Process Model and informed by psycholinguistic research using keystroke logging, this exploratory pilot study explores the potential effects of AI-generated feedback on EFL learners' writing quality and writing processes from both product- and process-oriented perspectives. Using a between-subjects experimental design, fourteen EFL undergraduates were randomly assigned to either a ChatGPT-assisted revision group or a self-revision control group. Participants produced an initial draft and subsequently revised their texts under their assigned condition. Writing processes were examined through Inputlog keystroke data capturing pausing behavior and revision patterns, while writing quality was assessed using an analytic rubric with established inter-rater reliability. The findings indicate that both revision conditions improved writing quality; however, participants who received AI-generated feedback achieved significantly higher post-revision writing quality scores than those who relied on self-revision. AI-assisted revision was also associated with reductions in several pausing- and revision-related measures, suggesting changes in observable writing behavior during revision. These findings highlight the potential of AI-generated feedback to support both the product and process dimensions of EFL writing.

Keywords: ChatGPT-assisted revision, writing fluency, writing quality, cognitive processes, keystroke logging

Copyright: © 2026 Oumaima Belgamra. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. **Data Availability Statement:** All relevant data are within this paper.

Introduction

The integration of Information and Communication Technology (ICT) into educational settings has increasingly become a focal point of research and pedagogical practice, as numerous studies underscore its potential to enhance teaching effectiveness, learning outcomes, and learner engagement. ICT encompasses a broad spectrum of digital tools, including computers, tablets, educational software, and internet-based platforms, which collectively facilitate knowledge acquisition, skill development, and active learning opportunities. In contemporary educational contexts, the use of such technologies by students has become ubiquitous, reflecting both global trends in digital literacy and the expanding role of technology in supporting formal and informal learning processes.

Generative AI Chatbots in Contemporary Educational Contexts

Within the broader landscape of information and communication technologies, the emergence of generative artificial intelligence tools such as ChatGPT has opened new and promising avenues for educational practice and research (Gayed et al., 2022; Mekheimer, 2025). Unlike traditional educational software, generative AI tools can provide readily available, interactive, and context-sensitive responses tailored to learners' queries and instructional needs. Among these technologies, AI-powered chatbots have gained particular prominence in educational settings, functioning as interactive conversational agents that facilitate communication, guide learners through both structured and open-ended tasks, and support knowledge construction and skill development.

Through continuous interaction, chatbots can create dynamic learning environments in which learners can ask questions, seek clarification, and practice skills through repeated, on-demand exchanges, thereby providing interactive support that may encourage learner engagement. Their ability to adapt feedback to learners' performance and preferences further supports personalized learning and has been shown to enhance motivation, engagement, and persistence across learning tasks (Benotti et al., 2017). Recent empirical research across diverse higher-education contexts has documented university students' use of and engagement with generative AI tools for academic and learning-related purposes (Kim et al., 2025; Kwakye et al., 2025; Machkour & Abriane, 2025; Stöhr et al., 2024).

Such engagement extends to language-learning contexts, where AI chatbots may support activities such as obtaining explanations of grammatical forms, exploring lexical choices, and receiving assistance with written production. In language learning settings, for instance, learners use these tools to develop metalinguistic awareness, clarify grammatical rules, expand lexical knowledge, and refine written production. Such patterns of use highlight the relevance of AI chatbots within contemporary learner-centered pedagogical approaches and underscore the need for empirical research that examines not only learning outcomes but also the processes through which these technologies shape learners' engagement and performance.

AI-Generated Feedback in EFL Writing

One of the most common ways learners currently engage with AI-based tools is through the use of chatbots such as ChatGPT to receive feedback on their written work and improve writing performance. Writing competence constitutes a central component of academic success, particularly for learners of English as a Foreign Language (EFL), for whom writing functions not only as a primary mode of assessment but also as a cognitive tool that supports language development across domains. Despite its importance, the development of effective writing skills in EFL contexts remains a complex and demanding process. Learners often have limited opportunities for sustained writing practice, while large class sizes and time constraints restrict instructors' capacity to provide frequent, detailed,

and individualized feedback. As a result, many EFL students experience persistent difficulties with grammatical accuracy, lexical selection, textual coherence, and overall writing quality, underscoring the need for instructional support mechanisms that can meaningfully facilitate writing development (Hyland, 2007; Storch, 2011).

In response to these challenges, technological advances have introduced increasingly accessible forms of writing support, most notably Automated Writing Evaluation systems and AI-based writing assistants such as ChatGPT. These tools enable learners to receive immediate, detailed, and context-sensitive feedback on multiple aspects of their writing, often accompanied by explanations that align with their level of understanding. AWE systems, in particular, rely on natural language processing techniques to analyze written texts across dimensions such as grammatical accuracy, lexical sophistication, coherence, organization, and overall quality (Ekizoğlu & Demir, 2025). Beyond the correction of surface-level errors, AI-assisted feedback may support more active engagement with the writing process by facilitating iterative revision and helping learners identify and address linguistic and organizational aspects of their texts (Mekheimer, 2025). Similarly, Teng's (2024b) systematic review identified both opportunities and challenges associated with ChatGPT use in EFL writing, highlighting its potential to provide immediate feedback and support writing development while also noting concerns related to AI dependency and the need for critical engagement with AI-generated output.

Theoretical Foundations of AI-Mediated Writing Feedback

The present study is primarily grounded in Flower and Hayes' (1981) Cognitive Process Model, which conceptualizes writing as a recursive cognitive activity involving three interrelated processes: planning, translating, and reviewing. By shifting attention from the final written product to the processes through which writers generate, organize, and refine ideas, the model provides a theoretical basis for examining both writing outcomes and the process-oriented measures considered in this study. Within this framework, AI-generated feedback can be conceptualized as external support operating particularly during reviewing and revision, where writers evaluate linguistic accuracy, coherence, and appropriateness while refining an evolving text. For EFL writers, these processes may be especially demanding, and AI-generated feedback may shape how learners identify and respond to linguistic and organizational issues during revision. Examining observable writing behavior alongside final-text quality therefore provides a means of considering AI-assisted writing from both process- and product-oriented perspectives.

While Flower and Hayes' (1981) Cognitive Process Model explains the cognitive processes involved in writing, Cognitive Load Theory (Sweller, 1988) offers a complementary perspective on the mental demands associated with these processes. According to Cognitive Load Theory, writing places substantial demands on working memory because writers must simultaneously plan ideas, retrieve lexical and grammatical knowledge, translate ideas into text, and monitor and revise their output. External instructional support may help learners manage these cognitive demands by providing guidance that enables them to allocate greater attentional resources to higher-order aspects of writing and revision. In the context of AI-assisted writing, feedback generated by tools such as ChatGPT may therefore assist learners in managing the cognitive demands associated with identifying and correcting linguistic problems during revision. Although the present study does not directly measure cognitive load, Cognitive Load Theory provides a complementary theoretical perspective for interpreting the observable changes in writing behavior identified in the present study.

Complementing this cognitive-process perspective, sociocognitive theories of learning conceptualize feedback as a form of scaffolding that can support learners within their zone of proximal development (Vygotsky, 1987). From this perspective, AI-generated feedback may provide temporary, task-specific

guidance that learners can draw on while revising their texts (Escalante et al., 2023; John-Steiner & Mahn, 1996). The extent to which such support contributes to learning, however, depends partly on how learners engage with the feedback: suggestions may be critically evaluated and selectively incorporated or adopted with limited reflection. This scaffolding perspective thus complements the Cognitive Process Model by accounting for the potential supportive role of external feedback within the writing and revision process.

Keystroke Logging as a Process-Tracing Method in Writing Research

Flower and Hayes' (1981) Cognitive Process Model conceptualizes writing as a dynamic and recursive activity in which writers continuously alternate between planning, translating, and reviewing. Crucially, these processes unfold over time and are not directly observable in the final written product. As a result, analyses that rely solely on text quality outcomes provide limited insight into how cognitive resources are allocated during writing or how instructional interventions reshape the writing process itself.

From a process-oriented perspective, temporal features of text production such as pausing, writing speed, and revision behavior provide observable indicators of how writing unfolds over time. Periods of sustained text production may be associated with relatively continuous text generation (Matsuhashi, 1981; Sasaki, 2000), whereas pauses and revisions may occur during activities such as planning, evaluation, or problem solving (Olive et al., 2009; Wallot & Grabowski, 2013). Importantly, pauses are not necessarily indicative of writing difficulty or reduced fluency, as they may accompany a range of writing-related activities, including planning, evaluation, and idea generation. Pause-related measures should therefore be interpreted as behavioral indicators of writing processes rather than as direct evidence of specific cognitive states.

To capture these dynamics empirically, the present study employs keystroke logging as a process-tracing method. Unlike traditional approaches such as think-aloud protocols or retrospective interviews, which rely on self-report and may interfere with natural writing behavior, keystroke logging enables the unobtrusive recording of writers' real-time activity. By logging every keystroke, pause, deletion, and revision, this method provides fine-grained temporal data that make it possible to examine writing fluency as it unfolds during text production.

Measures such as pause duration, burst characteristics, writing speed, and revision patterns provide process-level indicators of observable text-production behavior (Wengelin & Johansson, 2023). By examining these temporal and behavioral measures, the present study complements final-text assessment with a process-oriented examination of how writing and revision unfold under AI-assisted and self-revision conditions, in alignment with the process-oriented assumptions of Flower and Hayes' Cognitive Process Model.

Previous Research on AI-Mediated Writing Support

A growing body of research has examined AI-mediated feedback in writing, with findings varying across technologies, instructional contexts, and outcome measures. Studies focusing on writing and revision have documented benefits from different forms of AI support. Grammarly-based automated feedback, for example, was associated with improvements in EFL writing (Wei et al., 2023), while AI-assisted feedback produced greater writing gains than conventional teacher feedback in another instructional context (Ekizoğlu & Demir, 2025). Extending this evidence to generative AI, Meyer et al. (2024) reported better revision performance among students receiving LLM-generated feedback than among those revising without feedback. Research examining the feedback itself presents

a more complex picture: Dai et al. (2023) identified both correspondences and differences between ChatGPT-generated and instructor feedback, while Jacobsen and Weber (2025) demonstrated that the quality of LLM-generated feedback can vary according to factors such as prompting and the standard of comparison. Collectively, these findings indicate potential benefits for writing and revision while cautioning against treating AI-mediated feedback as a uniform intervention or assuming equivalence with human feedback across contexts.

Research has also extended beyond textual outcomes to cognitive, affective, and motivational dimensions. Essel et al. (2024) reported effects associated with ChatGPT use on critical, reflective, and creative thinking, whereas Huang and Teng (2025) observed significantly greater affective and behavioral engagement with ChatGPT feedback than with peer feedback, although the difference in cognitive engagement was not significant. Positive outcomes have additionally been reported for writing motivation, self-efficacy, engagement, and collaborative-writing orientation (Teng, 2024a), as well as dimensions of L2 motivation, particularly Ideal L2 Self and L2 Learning Experience (Huang & Mizumoto, 2024). Taken together, this strand of research suggests that AI-mediated support may be associated with outcomes extending beyond writing quality; however, variation in the constructs measured, interventions employed, and research designs limits broad generalization. Characteristics such as the immediacy and responsiveness of AI-mediated feedback may also shape learners' engagement with such support (Ene & Yao, 2021; Huang & Teng, 2025).

Recent research has further broadened the range of technologies and instructional configurations examined. Al-Sofi (2026) reported that Google Gemini identified common writing problems and was positively perceived by learners as a source of writing support, whereas Emerson et al. (2026) documented grammar gains following learner-initiated GenAI feedback in collaborative L2 writing that were maintained at a two-week delayed post-test. These studies extend the evidence base to newer forms of generative AI support, while differences in tools, tasks, feedback procedures, and research designs warrant caution when attributing observed outcomes specifically to AI or generalizing them across instructional contexts.

Taken together, existing research has generated considerable evidence concerning AI-mediated writing outcomes, feedback characteristics, learner perceptions, engagement, and motivation. Comparatively less attention, however, has been given to the real-time writing and revision processes associated with AI-generated feedback. In particular, process-oriented evidence derived from keystroke logging remains limited, leaving less understood how AI-assisted revision is reflected in observable temporal and behavioral patterns such as pausing, text production, and revision. Examining these process-level measures alongside writing quality may therefore complement existing product-based evidence and provide a more comprehensive account of AI-assisted revision.

Current Study

Addressing this process-oriented gap, the present exploratory pilot study combines final-text assessment with keystroke logging to examine AI-assisted and self-revision from complementary product- and process-oriented perspectives. Keystroke logging provides fine-grained temporal and behavioral data on text production, including pausing and revision patterns. Importantly, unlike methods such as think-aloud protocols or retrospective questionnaires, which rely on self-report and may interfere with or reconstruct aspects of the writing process, keystroke logging enables the non-intrusive recording of writing behavior as it unfolds. Its integration with final-text assessment therefore allows writing outcomes to be examined alongside observable patterns of text production and revision.

Given the exploratory nature of the investigation and the small sample, the study is intended as a pilot examination of AI-generated feedback in EFL writing rather than as a basis for broad generalization.

Accordingly, the study has two primary objectives. First, it explores potential differences in writing quality between AI-assisted revision and self-revision. Second, it examines whether observable writing-process patterns, including pausing and revision-related measures captured through keystroke logging, differ between the two revision conditions. By integrating product- and process-oriented measures, the study aims to provide a more comprehensive account of AI-assisted revision in EFL writing.

Compared with many previous studies that primarily evaluated final writing products or investigated rule-based automated writing evaluation (AWE) systems such as Grammarly, the present study examines both product and process dimensions of writing through keystroke logging and focuses specifically on ChatGPT, a generative AI system capable of providing context-sensitive and adaptive feedback. Drawing on both product- and process-based perspectives, the study aims to provide a more comprehensive understanding of how AI-generated feedback may influence EFL writing and extend current knowledge of AI-mediated writing support in higher education. Based on these objectives, the following research questions guide the investigation:

RQ1: To what extent does writing quality differ between EFL learners who revise their writing with AI-generated feedback and those who engage in self-revision?

RQ2: How do observable writing-process patterns, particularly pausing and revision-related measures, differ between AI-assisted revision and self-revision?

Methods

Participants

Seventeen first-year undergraduate students enrolled in the English Department at the Faculty of Education Sciences initially volunteered to participate in the study. Prior to the experimental writing task, all participants completed a five-minute baseline typing task to assess keyboarding proficiency and ensure that they were able to complete the computer-based writing tasks. During data screening conducted prior to statistical analysis, three participants were excluded because their keyboarding performance was markedly lower than that of the remaining participants and was considered likely to disproportionately influence the keystroke-based measures. The final analytic sample therefore consisted of fourteen participants, including eight participants in the ChatGPT-assisted revision group and six participants in the self-revision group.

All participants were non-native speakers of English with Arabic as their first language and were enrolled in the same academic program and year of study. They had successfully completed comparable English language and academic writing coursework and had been exposed to similar instructional content prior to participation. These characteristics were expected to reduce variability associated with educational background, writing experience, and language proficiency, thereby facilitating a more focused examination of revision processes under different feedback conditions.

Although no standardized English proficiency test was administered, participants were recruited from a relatively homogeneous educational and linguistic background. Nevertheless, the absence of a formal proficiency measure constitutes a limitation, and future research should incorporate standardized proficiency assessments to more precisely account for individual differences in language ability and responsiveness to AI-generated feedback. Prior to participation, all students provided informed consent. Ethical procedures were maintained throughout the study, including assurances of anonymity, the right to withdraw at any stage, and the secure management of keystroke data.

Design

This study employed a between-subjects experimental design to examine the effect of AI-generated feedback on EFL students' writing fluency and quality. Participants were randomly assigned to one of two conditions using a computer-generated online randomization procedure: an experimental group that received AI-generated feedback on their written texts and a control group that engaged in self-revision without external guidance. Writing performance was recorded using Inputlog keystroke logging software, enabling the analysis of process-oriented measures such as pausing behavior, revision activity, and production bursts. Writing quality was assessed using an analytic rubric. Together with the keystroke-logging data, these measures enabled the investigation of both product- and process-oriented dimensions of writing. A between-subjects design was adopted to minimize potential carryover effects that might arise if participants were exposed to both revision conditions, thereby preserving the independence of the experimental conditions.

To ensure the reliability and objectivity of writing quality assessment, all essays were independently evaluated by two adjunct professors of English at the Faculty of Education Sciences with experience in teaching and assessing EFL writing. Both raters used the analytic rubric developed by González et al. (2017), which evaluates content, organization, language use, vocabulary, and mechanics. The raters were blind to participants' condition and were not informed whether essays represented pre- or post-revision drafts. Inter-rater reliability was assessed using the average-measures Intraclass Correlation Coefficient (ICC) based on a two-way mixed-effects model with absolute agreement. Results indicated excellent agreement for both pre-revision texts (ICC = .898, 95% CI [.440, .972]) and post-revision texts (ICC = .965, 95% CI [.558, .992]). Although the average-measures ICCs indicated excellent inter-rater agreement, the relatively wide confidence intervals indicate some uncertainty around these reliability estimates because of the small sample size and should therefore be interpreted cautiously. Final writing quality scores were calculated by averaging the ratings assigned by the two evaluators.

Instruments

Inputlog 9.5.0.1

In the present study, writing fluency was assessed using Inputlog, a specialized keystroke logging tool designed to capture the dynamic processes involved in text production. Writing fluency encompasses both the speed and ease with which writers generate and organize text, reflecting the interaction between cognitive planning, language formulation, and motor execution. By recording detailed typing behavior, Inputlog allows researchers to quantify fluency in a precise and objective manner, going beyond traditional product-based measures of writing.

Inputlog systematically tracks each keystroke, along with the timing of inputs, pauses, deletions, and insertions, providing a continuous, fine-grained record of the writing process. Unlike traditional methods such as thinking-aloud protocols, Inputlog is less intrusive, allowing participants to write naturally without the cognitive burden of verbalizing their thoughts. This non-intrusive approach helps preserve ecological validity while capturing real-time behavioral patterns during writing.

For the purposes of this study, writing fluency was operationalized through several key process measures derived from Inputlog data:

- **Total pausing time:** the cumulative duration of all pauses, reflecting the cognitive effort involved in planning and problem-solving during writing.
- **Thinking-to-writing ratio:** the proportion of pause time relative to active typing time, indicating the balance between cognitive planning and text production.

- **Number of pauses:** the frequency of interruptions in typing, providing insight into moments of reflection, hesitation, or uncertainty.
- **Within-word pauses:** pauses occurring mid-word, signaling micro-level processing challenges or lexical retrieval difficulties.
- **Number of bursts:** sequences of uninterrupted keystrokes, representing periods of fluent text generation.
- **Deletions and insertions:** the frequency of text modifications, capturing revision behaviors and self-monitoring strategies.

Following common practice in keystroke logging research, pauses were operationalized as interruptions in keyboard activity lasting 2000 milliseconds (2 seconds) or longer. This threshold was used in Inputlog to distinguish meaningful pauses associated with planning, formulation, and revision processes from brief motor interruptions that occur during normal typing. Accordingly, all pause-related measures reported in the study, including total pausing time, total number of pauses, within-word pauses, and between-word pauses, were calculated using the 2000 ms threshold. Because pause-based measures can be influenced by keyboarding proficiency, participants completed a baseline typing task prior to the experimental writing task, and individuals exhibiting markedly lower typing performance were excluded during data screening to reduce variability associated with typing skill.

By employing Inputlog, the study was able to capture both temporal and structural aspects of writing fluency, offering a comprehensive view of how participants produce, monitor, and revise their written texts. This method thus provides valuable insights into observable writing processes and allows for a detailed comparison of fluency patterns across different experimental conditions.

The Writing Prompt

Participants were asked to compose a paragraph of 100–150 words addressing the topic: “*The Positive and Negative Effects of using Technology on Education.*” They were instructed to structure their writing according to standard academic conventions, beginning with a clear topic sentence, developing their ideas with coherent supporting explanations and illustrative examples, and concluding with a concise and appropriate closing sentence. This task was designed to elicit writing that reflects both organizational and linguistic competence, allowing for the assessment of content development, coherence, and argumentative clarity. The time limit for this task was 25 minutes.

All participants completed the same writing task, thereby standardizing the writing topic across the two revision conditions. The task addressed the positive and negative effects of technology in education and did not require specialized disciplinary knowledge. However, participants’ individual familiarity with or prior knowledge of the topic was not independently assessed. This is an important consideration, as topic familiarity has been shown to influence the allocation of time across L2 writing processes as well as text production and revision behavior (Li & Yu, 2024). Accordingly, individual differences in prior topic knowledge cannot be ruled out as a potential source of variability in the writing-process measures observed in the present study.

The Writing Analytic Rubric

The final written texts were evaluated using the analytic scoring rubric developed by González, Trejo, and Roux (2017), which draws on earlier frameworks (Jacobs, 1981) and is informed by the Common European Framework of Reference for Languages (CEFR). Rooted in the conceptualization of writing as a system of interrelated components (Weigle, 2002), this rubric is particularly well suited to EFL and ESL contexts (Hamp-Lyons, 1991; Weigle, 2002). It evaluates five key dimensions, namely

content, organization, language use, vocabulary, and mechanics/spelling, each rated on a 0–5-point scale, yielding a maximum score of 25. This analytic framework allows for a nuanced assessment by highlighting learners' specific strengths and areas for improvement. To ensure inter-rater reliability, all texts were independently scored by two university professors.

Procedure

The experimental procedure consisted of two sequential phases, designed to investigate the effects of AI-generated feedback on EFL learners' writing fluency and quality while maintaining controlled and replicable conditions.

Phase One: Initial Drafting

Prior to the experimental task, participants completed the standardized five-minute Inputlog Copy Task to assess keyboarding proficiency and familiarize themselves with the computer-based writing environment. This procedure was implemented because previous keystroke-logging research has shown that keyboarding proficiency may influence process measures such as pausing behavior and burst production (Leijten & Van Waes, 2013). Following the data-screening procedure described in the Participants section, three participants whose performance on the standardized copy task was markedly lower than that of the remaining participants were excluded from the final analyses, resulting in a final analytic sample of fourteen participants with relatively comparable keyboarding proficiency levels.

The typing task served as a screening procedure rather than as a statistical covariate; therefore, typing performance was not entered into the statistical analyses as a covariate. Instead, fluency measures were analyzed after excluding participants with markedly lower keyboarding proficiency. Following the baseline typing task, participants were asked to write a paragraph of 100–150 words on the topic: *"The Positive and Negative Effects of using Technology on Education."* They were instructed to construct their paragraphs with a clear topic sentence, supporting ideas and examples, and an appropriate concluding sentence in accordance with the Council of Europe Analytic Rubric. Participants were given twenty-five minutes to complete their initial drafts independently without external assistance.

The 25-minute time limit applied only to the initial drafting phase. This duration was selected to provide participants with sufficient time to generate ideas and complete their initial texts while minimizing the influence of time pressure on writing behavior, as the primary objective of the study was to examine revision processes rather than performance under strict time constraints. To ensure standardized writing conditions and comparability across participants, the same time limit was applied to all individuals, as allowing participants to complete the task at different times could have introduced additional variability unrelated to the revision conditions under investigation. The selected duration is also consistent with previous EFL paragraph-writing research, in which learners have commonly been allocated between 20 and 30 minutes to produce paragraphs of approximately 100–200 words (e.g., Apridayani et al., 2024; McDonough & Fuentes, 2015).

Phase Two: Revision

After completing the initial draft, participants entered the revision phase, with procedures differing according to group assignment. Participants in the experimental group used ChatGPT-5 to obtain AI-generated feedback on their drafts. They copied their paragraph into the system and submitted the following standardized prompt: *"Please review the following paragraph and provide feedback to improve it in terms of: 1) Content, 2) Organization, 3) Use of Language, 4) Use of Vocabulary, and 5) Mechanics and Spelling. Suggest specific edits or alternatives where necessary."* The feedback

included both sentence-level suggestions related to grammar, vocabulary, and mechanics, as well as broader comments concerning content and organization. To ensure consistency across participants, all members of the experimental group received feedback using the same prompt, and no follow-up interactions with ChatGPT were permitted. Participants were allowed to incorporate AI-generated suggestions into their revised drafts while remaining responsible for producing the final version of the text. The revision phase lasted twenty minutes.

The control group engaged in self-revision during the same twenty-minute period. Participants were instructed to revise their paragraphs independently while focusing on the same five dimensions addressed in the AI feedback: content, organization, language use, vocabulary, and mechanics/spelling. They were guided by the following instruction: *“Revise your paragraph focusing on the following five areas: 1) Content, 2) Organization, 3) Use of Language, 4) Use of Vocabulary, and 5) Mechanics and Spelling. Make any changes you think will improve your text.”* No AI or external assistance was provided, allowing participants to revise their texts using their own linguistic knowledge and judgment.

Throughout both phases, writing fluency was continuously monitored using Inputlog. Following the recommendations of Leijten and Van Waes (2013), the raw keystroke data were cleaned by removing logs unrelated to the writing task, such as those generated during the familiarization phase or participant registration. The cleaned dataset was subsequently analyzed to obtain process measures, including pause duration, pausing frequency, revision activity, burst production, insertions, deletions, and overall writing fluency.

Data Analysis

Statistical analyses were conducted using IBM SPSS Statistics (Version 26). Given the exploratory nature of the study, the relatively small sample size, and the distributional characteristics of the data, non-parametric statistical procedures were employed throughout. Baseline comparability between the ChatGPT-assisted revision and self-revision groups was first examined using a Mann–Whitney U test on the pre-revision writing quality scores. Between-group differences in post-revision writing quality and keystroke logging measures were subsequently analyzed using Mann–Whitney U tests, whereas within-group changes from pre- to post-revision were examined using Wilcoxon signed-rank tests. Exact two-tailed p values are reported throughout because of the small sample size. Effect sizes (r) were calculated as the absolute Z value divided by the square root of the total number of observations (N) and interpreted according to Cohen’s (1988) guidelines. Keystroke logging measures, including total pausing time, thinking-to-writing ratio, total number of pauses, within-word pauses, between-word pauses, P-bursts, insertions, and deletions, were analyzed separately to examine different dimensions of writing fluency and revision behavior. Given the exploratory nature of this pilot study, no adjustment for multiple statistical comparisons was applied; therefore, the findings should be interpreted with appropriate caution.

Results

RQ1: To what extent does writing quality differ between EFL learners who revise their writing with AI-generated feedback and those who engage in self-revision?

To provide a clearer representation of participants’ performance levels and score variability, descriptive statistics are presented in Tables 1 and 2. Table 1 reports the means, standard deviations, medians, and interquartile ranges for writing quality scores before and after revision, whereas Table 2 presents the median and interquartile range values for the keystroke logging measures.

Table 1. *Descriptive Statistics for Writing Quality Before and After Revision*

Group	Pre-revision Mean $\pm$ SD	Pre-revision Md (IQR)	Post-revision Mean $\pm$ SD	Post-revision Md (IQR)
Self-Revision	11.83 $\pm$ 3.66	11.5 (5.5)	15.00 $\pm$ 3.41	14.0 (6.0)
ChatGPT-Assisted Revision	10.88 $\pm$ 1.55	11.0 (2.0)	20.88 $\pm$ 4.39	23.0 (4.0)

Table 2. *Descriptive Statistics (Median and Interquartile Range) for Keystroke Logging Measures Before and After Revision*

Measure	Self-revision pre Md (IQR)	Self-revision post Md (IQR)	ChatGPT pre Md (IQR)	ChatGPT post Md (IQR)
Total Pausing Time	9.22 (5.75)	8.37 (2.65)	7.42 (5.01)	3.59 (1.71)
Thinking-Writing Ratio	49.34 (17.85)	54.13 (26.51)	45.20 (18.50)	33.71 (25.24)
Total Number of Pauses	118 (63)	71 (28)	77 (35)	33 (21)
Within-Word Pauses	14 (30.50)	9 (14.50)	14 (24.00)	3 (3.00)
Between-Word Pauses	49 (30.00)	24 (16.50)	34 (10.00)	4 (7.00)
P-Bursts	112 (65.00)	64 (26.50)	78 (37.00)	32 (18.00)
Insertions	6 (51.00)	27 (64.00)	8 (19.00)	3 (12.00)
Deletions	31 (44.50)	48 (24.00)	5 (13.00)	56 (69.00)

Note. Md = Median; IQR = Interquartile Range.

Before examining the effects of the revision conditions, baseline comparability between the two groups was assessed using a Mann–Whitney U test. The analysis revealed no statistically significant difference in pre-revision writing quality between the two groups ($U = 21.50$, $Z = -0.33$, exact $p = .755$). Although this finding provides no statistical evidence of baseline differences in writing quality, the small sample size precludes firm conclusions regarding baseline equivalence. To determine whether the magnitude of improvement differed between the two revision conditions, change scores (post-revision minus pre-revision writing quality) were subsequently compared using a Mann–Whitney U test. The analysis revealed a statistically significant difference between the ChatGPT-assisted revision group and the self-revision group ($U = 7.00$, $Z = -2.22$, exact $p = .029$, $r = .59$), indicating that participants who received ChatGPT-assisted feedback demonstrated significantly greater improvements in writing quality than those who engaged in self-revision. The descriptive statistics presented in Tables 1 and 2 provide an overview of participants' performance and revision behaviors, while the statistical significance and magnitude of the observed differences are examined through the non-parametric analyses reported below.

The descriptive statistics suggest that both ChatGPT-assisted revision and self-revision were associated with improvements in writing quality, although the magnitude and consistency of these gains differed between the two revision conditions. Consistent with the significant difference observed in the change-score analysis, descriptive statistics showed higher post-revision writing quality in the ChatGPT-assisted revision group than in the self-revision group. A Mann–Whitney U test (Table 3) further indicated that the experimental group achieved a higher mean rank ($M = 9.63$, Sum of Ranks = 77.00) than the control group ($M = 4.67$, Sum of Ranks = 28.00). This difference was statistically significant ($U = 7.00$, $Z = -2.22$, exact $p = .029$) and associated with a large effect size ($r = .59$), indicating that ChatGPT-assisted revision resulted in significantly greater improvements in writing quality than self-revision.

Table 3. Mann–Whitney *U* Test Comparing Writing Quality Between AI-Feedback and Self-Revision Groups

Revision type	N	Mean rank	Sum of ranks	U	Z	Exact sig. (2-tailed)	<i>r</i>
Self-revision	6	4.67	28.00	07.00	−2.22	.029	.59
AI-Feedback	8	9.63	77.00				

Within-group analyses further clarify these findings. In the experimental group, all eight participants improved their writing quality following ChatGPT-assisted revision, as reflected by eight positive ranks and no negative ranks (Table 4). The Wilcoxon signed-rank test confirmed a significant increase in writing quality ($Z = -2.52$, exact $p = .008$) with a very large effect size ($r = .89$), indicating that the improvement was observed consistently across all participants. Participants in the self-revision group also demonstrated significant gains, with all six participants showing positive ranks and no negative ranks (Table 5). The Wilcoxon signed-rank test likewise revealed a significant improvement ($Z = -2.23$, exact $p = .031$) with a very large effect size ($r = .91$). However, the magnitude of improvement remained smaller than that observed in the ChatGPT-assisted revision group.

Taken together, these findings indicate that both revision conditions were associated with improved writing quality; however, participants who received ChatGPT-assisted feedback demonstrated significantly greater gains than those who relied on self-revision. These results suggest that AI-generated feedback may provide additional support during the revision process, enabling EFL learners to produce higher-quality revised texts.

Table 4. Wilcoxon Signed-Rank Test Comparing Writing Quality Before and After ChatGPT Feedback

	N	Mean rank	Sum of ranks	Z	Exact sig. (2-tailed)	<i>r</i>
Before Revision	0 ^a	.00	.00	−2.52	.008	.89
After Revision	8 ^b	4.50	36.00			

Table 5. Wilcoxon Signed-Rank Test Comparing Writing Quality Before and After Self-Revision

	N	Mean rank	Sum of ranks	Z	Exact sig. (2-tailed)	<i>r</i>
Before Revision	0 ^a	.00	.00	−2.23	.031	0.91
After Revision	6 ^b	3.50	21.00			

RQ2: How do observable writing-process patterns, particularly pausing and revision-related measures, differ between AI-assisted revision and self-revision?

The results suggest a clear and consistent pattern: AI-generated feedback was associated with substantial changes in observable writing behavior, whereas self-revision produced comparatively modest changes. Participants who received AI feedback exhibited fewer pauses but greater deletion activity during revision, together with reductions in both within- and between-word pauses. By contrast, self-revision produced relatively limited changes across the keystroke measures.

The large effect sizes observed in the present study should be interpreted with caution. Effect size estimates may be inflated in studies involving small and relatively homogeneous samples, particularly in pilot research. Consequently, although the findings suggest potentially meaningful effects, replication with larger and more diverse samples is necessary to establish the stability and generalizability of the observed effect sizes.

When comparing groups (Table 6), participants in the AI-feedback condition showed significantly lower total pausing time ($U = 2.00$, $Z = -2.84$, Exact $p = .003$, $r = .76$), fewer total pauses ($U = 3.00$, $Z = -2.71$, Exact $p = .005$, $r = .72$), fewer within-word pauses ($U = 0.00$, $Z = -3.01$, Exact $p = .001$, $r = .80$), fewer between-word pauses ($U = 3.50$, $Z = -2.51$, Exact $p = .008$, $r = .67$), and fewer P-bursts ($U = 1.00$, $Z = -2.97$, Exact $p = .001$, $r = .79$). These large effect sizes are consistent with substantial differences between the two revision conditions, although they should be interpreted cautiously given the exploratory nature of the study. In contrast, thinking-writing ratio, total insertions, and total deletions did not differ significantly between groups, suggesting that AI feedback primarily influenced pausing behavior and observable revision behavior.

Within-group analyses further clarify these patterns. Following ChatGPT-assisted revision (Table 7), participants significantly reduced total pausing time ($Z = -2.52$, Exact $p = .008$, $r = .89$), total number of pauses ($Z = -2.52$, Exact $p = .008$, $r = .89$), within-word pauses ($Z = -2.38$, Exact $p = .016$, $r = .84$),

Table 6. Mann–Whitney U Test Comparing Writing Fluency Between AI-Feedback and Self-Revision Groups

	Mann Whitney U	Wilcoxon W	Z	Exact sig. (2-tailed)	r
Total Pausing Time	2.00	38.00	-2.84	.003	0.76
Thinking Writing Ratio	9.00	45.00	-1.93	.059	0.52
Total Number of Pauses	3.00	39.00	-2.71	.005	0.72
Within Words Pauses	.00	28.00	-3.01	.001	0.80
Between Words Pauses	3.50	31.50	-2.51	.008	0.67
Number of P Bursts	1.00	37.00	-2.97	.001	0.79
Total Number of Insertions	9.50	45.50	-1.89	.059	0.51
Total Number of Deletions	14.00	35.00	-1.29	.228	0.34

Table 7. Wilcoxon Signed-Rank Test Comparing Writing Fluency Before and After ChatGPT Feedback

	Z	Exact sig. (2-tailed)	r
Total Pausing Time	-2.52	.008	0.89
Thinking Writing Ratio	-1.54	.141	0.55
Total Number of Pauses	-2.52	.008	0.89
Within Words Pauses	-2.38	.016	0.84
Between Words Pauses	-2.36	.016	0.84
Number of P Bursts	-2.52	.008	0.89
Total Number of Insertions	-.91	.406	0.32
Total Number of Deletions	-2.38	.016	0.84

between-word pauses ($Z = -2.36$, Exact $p = .016$, $r = .84$), and number of P-bursts ($Z = -2.52$, Exact $p = .008$, $r = .89$), while exhibiting a significant increase in deletion activity ($Z = -2.38$, Exact $p = .016$, $r = .84$). These very large effect sizes indicate substantial changes in observable writing behavior following AI-assisted revision. Conversely, thinking-writing ratio ($Z = -1.54$, Exact $p = .141$, $r = .55$) and total insertions ($Z = -0.91$, Exact $p = .406$, $r = .32$) did not change significantly.

By contrast, self-revision (Table 8) produced more limited changes. Participants exhibited reductions in within-word pauses ($Z = -1.99$, Exact $p = .063$, $r = .81$) and between-word pauses ($Z = -1.99$, Exact $p = .063$, $r = .81$); however, these changes did not reach statistical significance. Likewise, total pausing time ($Z = -1.36$, Exact $p = .219$, $r = .56$), thinking-writing ratio ($Z = -0.94$, Exact $p = .430$, $r = .38$), total number of pauses ($Z = -1.57$, Exact $p = .150$, $r = .64$), number of P-bursts ($Z = -1.78$, Exact $p = .094$, $r = .73$), total insertions ($Z = -0.36$, Exact $p = .875$, $r = .15$), and total deletions ($Z = -0.94$, Exact $p = .438$, $r = .38$) also remained statistically unchanged. Overall, these findings suggest that self-revision resulted in relatively limited changes in writing fluency.

Table 8. *Wilcoxon Signed-Rank Test Comparing Writing Fluency Before and After Self-Revision*

	Z	Exact sig. (2-tailed)	r
Total Pausing Time	-1.36	.219	0.56
Thinking Writing Ratio	-.94	.43	0.38
Total Number of Pauses	-1.57	.15	0.64
Within Words Pauses	-1.99	.063	0.81
Between Words Pauses	-1.99	.063	0.81
Number of P Bursts	-1.78	.094	0.73
Total Number of Insertions	-.36	.875	0.15
Total Number of Deletions	-.94	.438	0.38

Taken together, the findings indicate that AI-generated feedback was associated with broader changes in observable writing behavior than self-revision. Across multiple measures, participants who received AI feedback exhibited reductions in several pausing measures together with increased deletion activity, whereas self-revision produced comparatively limited changes across the keystroke measures. These results suggest that AI-generated feedback may be associated with different revision processes than self-directed revision alone.

Discussion

The present study examined the effects of AI-generated feedback, specifically ChatGPT, on EFL undergraduate students' writing quality and writing fluency, comparing these effects with those of self-revision. By adopting both product-oriented and process-oriented measures, the study provides a nuanced account of how AI-mediated feedback influences not only what learners write but also how they write. The findings reveal that AI feedback exerts differential yet complementary effects on writing quality and writing fluency, underscoring the importance of treating these constructs as theoretically and empirically distinct.

Effect of AI-Generated Feedback on Writing Quality

With respect to writing quality, the findings indicate that both ChatGPT-assisted revision and self-revision resulted in statistically significant improvements across drafting sessions. This finding is

consistent with the well-established view that revision plays a crucial role in enhancing written performance. However, the experimental group that received AI-generated feedback significantly outperformed the self-revision group in post-revision writing quality, suggesting that AI-mediated feedback may provide additional pedagogical value beyond learners' independent revision efforts.

The greater improvement in writing quality following AI-assisted revision may be attributable to the explicit, immediate, and multidimensional nature of the feedback provided. Unlike self-revision, which depends largely on learners' existing linguistic knowledge and metalinguistic awareness, ChatGPT delivers targeted guidance addressing multiple dimensions of writing, including grammar, vocabulary, coherence, organization, and mechanics. Such feedback may facilitate linguistic noticing (Schmidt, 1990) by drawing learners' attention to discrepancies between their interlanguage and target-like forms, thereby supporting more effective revision decisions. This interpretation aligns with previous research demonstrating that AI-based feedback systems can promote engagement with both language form and textual organization, leading to higher-quality written outcomes (Ekizoğlu & Demir, 2025; Wei et al., 2023; Dai et al., 2023).

These findings are also broadly consistent with previous empirical studies. For instance, the quasi-experimental study by Ekizoğlu and Demir (2025) reported significant improvements in students' overall writing performance following AI-assisted instruction. Similarly, previous research has highlighted the potential of AI-generated feedback to improve writing quality, particularly in contexts where individualized teacher feedback is limited (Nazari et al., 2021; Ngo et al., 2024). The present findings extend this body of research by suggesting that these benefits may persist even when AI-assisted revision is compared directly with self-revision.

The consistency of improvement observed among participants in the experimental group suggests that ChatGPT functioned as an effective source of external support during revision. By providing accessible feedback across multiple dimensions of writing, the tool may have helped learners identify and address issues that might otherwise have gone unnoticed during self-revision. While self-revision remains valuable for fostering autonomy and reflective skills, its effectiveness may be constrained by learners' linguistic knowledge and ability to detect weaknesses in their own texts, which may partly explain the comparatively smaller gains observed in the control group.

Effect of AI-Generated Feedback on Writing Fluency

In contrast to writing quality, which reflects the linguistic and rhetorical characteristics of the final written product, writing fluency captures the writing process, including the temporal, procedural, and behavioral dynamics of text production as it unfolds in real time. The findings of the present study demonstrate that ChatGPT-assisted revision exerted a pronounced and systematic effect on writing fluency. Participants who received AI-generated feedback exhibited significant reductions in total pausing time, overall frequency of pauses, within-word and between-word pauses, and P-bursts, together with a significant increase in deletion activity. Collectively, these findings suggest substantial changes in observable writing behavior during revision. However, because pausing measures and P-burst counts may also be influenced by factors such as text length, active writing time, and revision strategies, they should be interpreted as behavioral indicators of writing processes rather than direct evidence of improved writing fluency. By comparison, self-revision produced only limited fluency-related gains, primarily at the micro-level of pausing behavior, with no significant improvements observed in broader fluency measures such as total pausing time or revision intensity.

From a psycholinguistic perspective, these findings may be interpreted through the lens of Cognitive Load Theory (Sweller, 1988) and the Cognitive Process Model (Flower & Hayes, 1981). Writing is

a cognitively demanding activity that requires the simultaneous coordination of planning, idea generation, lexical retrieval, syntactic encoding, and self-monitoring. Frequent pausing and deletions are often interpreted as indicators of processing difficulty, uncertainty, or disruptions in lexical and syntactic formulation (Sinharay et al., 2019). Consequently, the substantial reduction in pausing behavior together with the increase in deletion activity following AI-assisted revision may suggest that participants revised their texts differently while spending less time pausing during the revision process. These findings may therefore be consistent with reduced attentional demands during revision and greater procedural stability during writing.

However, because keystroke logging provides only indirect behavioral indicators, these findings should not be interpreted as direct evidence of specific cognitive mechanisms. Alternative explanations remain plausible. For example, some learners may have incorporated AI-generated suggestions with limited critical evaluation, resulting in cognitive offloading or passive transcription rather than deeper linguistic processing. Consequently, the present study cannot determine whether AI feedback primarily facilitated language learning or merely supported more efficient task completion. Nevertheless, keystroke logging remains valuable because it captures process-level behaviors that are not evident from analyses of the final written product alone. By recording every keystroke, pause, and revision in real time, Inputlog provided detailed evidence of how AI-assisted revision influenced writing processes throughout the revision stage, offering insights that could not have been obtained through product-based assessment alone.

An interesting finding concerns the self-revision group, which demonstrated a significant improvement in writing quality despite relatively limited changes in fluency measures. This pattern may suggest that improvements in text quality do not necessarily require substantial changes in observable writing behavior. Participants may have enhanced their texts through focused revisions, such as correcting grammatical errors, refining lexical choices, improving sentence clarity, and resolving local textual problems. Such targeted modifications can improve overall writing quality without necessarily producing large changes in insertion, deletion, or pausing measures. Furthermore, although self-revision improved writing quality, the final writing scores remained considerably lower than those achieved following ChatGPT-assisted revision. This suggests that self-revision primarily supported localized improvements to the existing text, whereas AI-generated feedback was associated with substantially higher-quality final drafts. Although reductions in within-word and between-word pauses were observed in the self-revision group, these changes did not reach statistical significance when exact significance values were calculated. Therefore, no firm conclusions can be drawn regarding underlying cognitive changes associated with these measures.

An important distinction should be made between immediate performance enhancement and long-term language acquisition. The present findings demonstrate that ChatGPT-assisted revision was associated with improvements in writing quality and observable writing behavior during a single revision session. However, because the study did not include a delayed post-test, it remains unclear whether these improvements reflect durable learning or temporary performance gains supported by external assistance. One possible interpretation is that AI-generated feedback functions as a pedagogical scaffold, helping learners notice linguistic problems and gradually internalize effective revision strategies. An alternative interpretation is that AI feedback primarily serves as a temporary support, enabling learners to produce stronger texts without necessarily developing the underlying knowledge and skills required for independent writing. Consequently, the present study cannot determine whether the observed benefits reflect genuine language development, improved task performance, or a combination of both. Longitudinal research incorporating delayed post-tests is therefore needed to determine whether the benefits associated with AI-assisted revision are sustained once external support is removed.

Synthesis and Theoretical Implications

Taken together, the findings suggest that AI-generated feedback may support writing development at both the product and process levels. Participants who received ChatGPT-assisted feedback produced higher-quality texts while also demonstrating changes in observable writing behaviors, including reductions in pausing together with increased deletion activity. These findings highlight the potential of AI-generated feedback to influence not only the final written product but also the way revision is carried out during writing. From a theoretical perspective, the results are broadly consistent with models of writing that emphasize the interaction between planning, text production, and revision processes. While self-revision remains valuable for fostering learner autonomy and reflective practice, AI-mediated feedback may serve as a complementary source of support for EFL learners engaged in cognitively demanding writing tasks.

Concluding Remarks

This exploratory pilot study contributes to the emerging literature on AI-assisted writing by integrating product- and process-oriented perspectives through the use of keystroke logging. The findings suggest that ChatGPT-assisted revision was associated with greater improvements in writing quality and broader changes in observable writing behavior than self-revision, highlighting the potential of AI-generated feedback to support both writing outcomes and revision processes in EFL contexts. However, because keystroke measures provide indirect behavioral indicators and the study involved a relatively small sample and a single revision session, the findings should be interpreted as preliminary rather than definitive. Future research employing larger and more diverse samples, longitudinal designs, delayed post-tests, and complementary qualitative methods is needed to determine whether the observed process-level changes translate into sustained improvements in independent writing performance.

Pedagogical Implications

The findings of this study offer several practical implications for the teaching of academic writing in EFL contexts. One particularly promising application involves the use of AI-generated feedback as part of a structured revision process. Students may first use AI tools such as ChatGPT to identify and address linguistic issues related to grammar, vocabulary, mechanics, and clarity, allowing them to refine initial drafts before receiving teacher feedback. This approach may enable instructors to devote more attention to higher-order aspects of writing, such as argumentation, organization, originality, and critical engagement, thereby maximizing the complementary strengths of both AI and human feedback. Such an approach may be particularly valuable in EFL contexts characterized by large class sizes and limited opportunities for individualized feedback.

At the same time, the effective integration of AI-assisted revision requires the development of AI literacy skills. Students should be encouraged to formulate effective prompts and critically evaluate AI-generated suggestions rather than accepting them uncritically, as AI systems may occasionally provide inaccurate, incomplete, or contextually inappropriate feedback. The quality of AI-generated output depends partly on the quality of the prompts provided and the user's ability to assess the appropriateness of the responses. Consequently, AI-generated feedback should be viewed as a support tool that complements, rather than replaces, learners' own judgment and teacher guidance. This emphasis on AI literacy is consistent with recent research showing that learners' effective use of AI tools depends not only on access to the technology itself but also on their ability to evaluate, interpret, and appropriately integrate AI-generated support into their writing processes (Tran, 2026).

A second pedagogical implication concerns the potential use of keystroke logging as a metacognitive tool. By providing learners with visual representations of their writing processes, including pausing patterns, revision behaviors, and production bursts, keystroke logging may encourage greater awareness of individual writing strategies and areas for improvement. Students can use these visualizations to identify moments where writing fluency breaks down, reflect on revision behaviors, and develop more effective writing strategies. When combined with AI-generated feedback and teacher guidance, keystroke logging may support a more reflective, process-oriented approach to writing instruction that promotes writing development, learner self-regulation, and metacognitive awareness.

Limitations and Future Research

Despite its contributions, this exploratory pilot study has several limitations that should be considered when interpreting the findings. First, the study relied on a relatively small and homogeneous sample. Given the time-intensive nature of keystroke logging data collection and analysis, the sample size was determined by participant availability rather than a priori statistical power considerations. As a result, the findings should be interpreted as exploratory because small samples may limit statistical generalizability, increase sensitivity to individual variability, and produce inflated effect size estimates. Furthermore, the relatively large number of statistical comparisons conducted in the present study increases the possibility of Type I error. Replication with larger and more diverse learner populations is therefore necessary to establish the robustness and generalizability of the findings.

Second, although participants were recruited from the same academic program and shared a relatively homogeneous educational and linguistic background, no standardized measure of English language proficiency was administered. In addition, participants were excluded following post hoc screening of markedly atypical keyboarding performance rather than on the basis of a prespecified numerical typing-speed criterion. Although the baseline typing task helped reduce variability associated with keyboarding proficiency, some degree of selection bias cannot be ruled out. Future studies should incorporate standardized measures of English proficiency, establish objective keyboarding-proficiency criteria prior to data collection, and statistically account for individual differences such as prior experience with AI tools, self-regulation skills, and typing ability.

A further limitation concerns participants' familiarity with the writing topic. Although all participants completed the same writing task, participants' individual familiarity with or prior knowledge of the topic was not independently assessed. Differences in prior topic knowledge may therefore have introduced additional variability into the writing-process measures. Future research should directly assess topic familiarity prior to task completion or systematically manipulate familiarity across writing tasks.

Another limitation concerns the comparison between AI-assisted revision and self-revision. Although both groups revised their texts according to the same writing dimensions, only the AI-assisted group received individualized and elaborated feedback. Moreover, participants in the AI-assisted condition spent part of the revision phase reading and processing ChatGPT-generated feedback before revising their texts, whereas participants in the self-revision condition revised without external input. As a result, differences in keystroke patterns may partly reflect differences in task structure rather than the effects of AI-generated feedback alone. Accordingly, some of the observed advantages may be attributable to the benefits of structured external guidance rather than the unique effects of AI-generated feedback itself. Future research should therefore compare AI-generated feedback with other forms of structured support, such as teacher feedback, peer feedback, or checklist-guided revision, to better isolate the specific contribution of AI-assisted revision. The present findings should thus be interpreted as demonstrating the effectiveness of AI-assisted revision relative to self-revision under the conditions of this study rather than as evidence that AI feedback is superior to all alternative forms of feedback.

An additional limitation concerns the interpretation of writing processes. Although keystroke logging provides detailed information about pausing behavior, revision activity, and other observable aspects of writing, it cannot reveal the qualitative nature of textual changes or the cognitive processes underlying these behaviors. Consequently, the present study cannot determine whether revisions primarily involved surface-level modifications or higher-order changes related to content, organization, and meaning. Future research could combine keystroke logging with qualitative analyses of revision types, such as the taxonomy proposed by Faigley and Witte (1981), to provide a more comprehensive understanding of how AI-generated feedback shapes revision processes.

Finally, the study examined only the immediate effects of AI-generated feedback during a single revision session using one writing task and a single standardized ChatGPT prompt. Consequently, it remains unclear whether the observed improvements reflect durable learning or temporary performance gains supported by external assistance. Although AI-generated feedback may facilitate revision during individual writing tasks, its long-term influence on learners' independent writing development remains uncertain. Longitudinal studies incorporating multiple writing tasks and delayed post-tests are needed to determine whether the observed improvements are maintained over time after external support is withdrawn.

Future research may also explore the pedagogical potential of keystroke logging as a metacognitive tool. Investigating whether regular exposure to keystroke visualizations, reflective prompts, and strategy instruction enhances learners' self-regulation, writing fluency, and writing quality would further extend the process-oriented contributions of the present study.

Acknowledgment

This work was carried out with the support of the Centre National de la Recherche Scientifique et Technique (CNRST) as part of the 'PhD-Associate Scholarship – PASS' program. I would also like to express my sincere appreciation to the editors and anonymous reviewers for their thoughtful and constructive feedback, which greatly contributed to strengthening the manuscript.

References

- Al-Sofi, B. (2026). The effectiveness of AI-generated feedback in enhancing students' writing proficiency: A study of Google Gemini. *Technology in Language Teaching & Learning*, 8, 103403. <https://doi.org/10.29140/tltl.2026.103403>
- Apridayani, A., Han, W., & Sakkanayok, K. (2024). Enhancing English writing competence in higher education: A comparative study of teacher-only assessment versus teacher and student self-assessment approaches. *Asian-Pacific Journal of Second and Foreign Language Education*, 9(1), 37. <https://doi.org/10.1186/s40862-024-00263-3>
- Benotti, L., Martinez, M. C., & Schapachnik, F. (2017). A tool for introducing computer science with automatic formative assessment. *IEEE Transactions on Learning Technologies*, 11(2), 179–192. <https://doi.org/10.1109/TLT.2017.2682084>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
- Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y.-S., Gašević, D., & Chen, G. (2023). Can large language models provide feedback to students? A case study on ChatGPT. In *2023 IEEE international conference on advanced learning technologies (ICALT)* (pp. 323–325). IEEE. <https://doi.org/10.1109/ICALT58122.2023.00100>
- Ekizoğlu, M., & Demir, A. N. (2025). The role of AI-assisted writing feedback in developing secondary students' writing skills. *Discover Education*, 4, 454. <https://doi.org/10.1007/s44217-025-00919-3>

- Emerson, N., Holland, T., Mitchell, T. D., Shum, D., & Oga-Baldwin, W. Q. (2026). Embedding constrained, learner-initiated GenAI feedback in collaborative L2 writing: Targeted grammar gains with a two-week follow-up. *Technology in Language Teaching & Learning*, 8. <https://doi.org/10.29140/tltl.2026.104008>
- Ene, E., & Yao, J. (2021). How does that make you feel: Students' affective engagement with feedback. *Language Teaching Research Quarterly*, 25, 66–83. <https://doi.org/10.32038/ltrq.2021.25.04>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, 6, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
- Faigley, L., & Witte, S. (1981). Analyzing revision. *College Composition and Communication*, 32(4), 400–414. <https://doi.org/10.2307/356602>
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387. <https://doi.org/10.2307/356600>
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- González, E. F., Trejo, N. P., & Roux, R. (2017). Assessing EFL university students' writing: A study of score reliability. *Revista Electrónica de Investigación Educativa*, 19(2), 91–103. <https://doi.org/10.24320/redie.2017.19.2.928>
- Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241–276). Ablex.
- Huang, J., & Mizumoto, A. (2024). Examining the effect of generative AI on students' motivation and writing self-efficacy. *Digital Applied Linguistics*, 1. <https://doi.org/10.29140/dal.v1.102324>
- Huang, J., & Teng, M. F. (2025). Peer feedback and ChatGPT-generated feedback on Japanese EFL students' engagement in a foreign language writing context. *Digital Applied Linguistics*, 2. <https://doi.org/10.29140/dal.v2.102469>
- Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of Second Language Writing*, 16(3), 148–164. <https://doi.org/10.1016/j.jslw.2007.07.005>
- Jacobs, H. L. (1981). *Testing ESL composition: A practical approach*. Newbury House Publishers.
- Jacobsen, L. J., & Weber, K. E. (2025). The promises and pitfalls of large language models as feedback providers: A study of prompt engineering and the quality of AI-driven feedback. *AI*, 6(2), 35. <https://doi.org/10.3390/ai6020035>
- John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist*, 31(3–4), 191–206. https://doi.org/10.1207/s15326985ep3103&4_4
- Kim, J., Klopfer, M., Grohs, J. R., Eldardiry, H., Weichert, J., Cox, L. A., II, & Pike, D. (2025). Examining faculty and student perceptions of generative AI in university courses. *Innovative Higher Education*, 50, 1281–1313. <https://doi.org/10.1007/s10755-024-09774-w>
- Kwakye, I. N., Kwakye, K. A. P., Adom-Fynn, D., Mensah, F. E., & Agyei, S. D. (2025). Exploring the use of generative artificial intelligence tools among university students in Ghana. *Discover Education*, 4, 432. <https://doi.org/10.1007/s44217-025-00608-1>
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30(3), 358–392. <https://doi.org/10.1177/0741088313491692>
- Li, S., & Yu, H. (2024). Effects of topic familiarity on L2 writing processes and behaviors. *International Journal of Applied Linguistics*, 34(1), 348–366. <https://doi.org/10.1111/ijal.12497>

- Machkour, B., & Abriane, A. (2025). Exploring students' attitudes toward generative artificial intelligence through the use of ChatGPT: A qualitative study in higher education. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 14, e33461. <https://doi.org/10.14201/adcaij.33461>
- Matsuhashi, A. (1981). Pausing and planning: The tempo of written discourse production. *Research in the Teaching of English*, 15(2), 113–134. <https://doi.org/10.58680/rte198115773>
- McDonough, K., & Fuentes, C. G. (2015). The effect of writing task and task conditions on Colombian EFL learners' language use. *TESL Canada Journal*, 32(2), 67–79. <https://doi.org/10.18806/tesl.v32i2.1208>
- Mekheimer, M. (2025). Challenges in the preparation of EFL students in faculties of education in Egypt. *Social Sciences & Humanities Open*, 11, 101517. <https://doi.org/10.1016/j.ssaho.2025.101517>
- Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, 6, 100199. <https://doi.org/10.1016/j.caeai.2023.100199>
- Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of artificial intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- Ngo, T. T. N., Chen, H. H. J., & Lai, K. K. W. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: A three-level meta-analysis. *Interactive Learning Environments*, 32(2), 727–744. <https://doi.org/10.1080/10494820.2022.2096642>
- Olive, T., Alves, R. A., & Castro, S. L. (2009). Cognitive processes in writing during pause and execution periods. *European Journal of Cognitive Psychology*, 21(5), 758–785. <https://doi.org/10.1080/09541440802079850>
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, 9(3), 259–291. [https://doi.org/10.1016/S1060-3743\(00\)00028-X](https://doi.org/10.1016/S1060-3743(00)00028-X)
- Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Sinharay, S., Zhang, M., & Deane, P. (2019). Prediction of essay scores from writing process and product features using data mining methods. *Applied Measurement in Education*, 32(2), 116–137. <https://doi.org/10.1080/08957347.2019.1577245>
- Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers and Education: Artificial Intelligence*, 7, 100259. <https://doi.org/10.1016/j.caeai.2024.100259>
- Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, 31, 275–288. <https://doi.org/10.1017/S0267190511000079>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Teng, M. F. (2024a). “ChatGPT is the companion, not enemies”: EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- Teng, M. F. (2024b). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies*, 6(3). <https://doi.org/10.58304/ijts.20240304>
- Tran, H. (2026). How does AI work as a “game changer”? Self-determination and influencing factors in EFL students' use of AI for academic writing. *Technology in Language Teaching & Learning*, 8. <https://doi.org/10.29140/tltl.2026.103723>
- Vygotsky, L. S. (1987). Thinking and speech (N. Minick, Trans.). In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky: Vol. 1. Problems of general psychology* (pp. 39–285). Plenum Press. (Original work published 1934)

- Wallot, S., & Grabowski, J. (2013). Typewriting dynamics: What distinguishes simple from complex writing tasks? *Ecological Psychology*, 25(3), 267–280. <https://doi.org/10.1080/10407413.2013.810512>
- Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: A randomized controlled trial. *Frontiers in Psychology*, 14, 1249991. <https://doi.org/10.3389/fpsyg.2023.1249991>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732997>
- Wengelin, Å., & Johansson, V. (2023). Investigating writing processes with keystroke logging. In O. Kruse, C. Rapp, C. M. Anson, K. Benetos, E. Cotos, A. Devitt, & A. Shibani (Eds.), *Digital writing technologies in higher education: Theory, research, and practice* (pp. 405–420). Springer. https://doi.org/10.1007/978-3-031-36033-6_25