



AI-Generated Feedback and Second Language Writing Development: A Longitudinal Investigation of Learner Autonomy, Accuracy, and Affective Response in Higher Education Contexts

Mohammed Khalifa Joghdam *

Department of English, Faculty of Arts and Languages, University of Tripoli, Tripoli, Libya

الملاحظات المولدة بالذكاء الاصطناعي وتطوير كتابة اللغة الثانية:
تحقيق تنبعي في الاستقلالية لدى المتعلمين، والدقة، والاستجابة الوجدانية في سياق
التعليم العالي

محمد خليفة جغدام *

قسم اللغة الإنجليزية، كلية الآداب واللغات، جامعة طرابلس، طرابلس، ليبيا

*Corresponding author: mkjoghdam@gmail.com

Received: April 11, 2026

Accepted: June 19, 2026

Published: June 25, 2026

Abstract:

This longitudinal mixed-methods study investigates the long-term impact of AI-generated feedback tools such as ChatGPT and Grammarly on undergraduate EFL/ESL learners' academic writing development over three academic years[1][2]. Addressing gaps in existing automated written corrective feedback (AWCF) research, the study examines three key dimensions: linguistic development, learner autonomy, and affective outcomes. Approximately 80 university students will be tracked across six semesters through quantitative measures of syntactic complexity, lexical diversity, error frequency, writing anxiety, self-efficacy, and foreign language enjoyment, alongside qualitative data from interviews, reflective journals, stimulated recall sessions, and screen-capture logs. The study aims to provide deeper insight into how learners engage with AI-generated feedback and how it influences their revision practices, long-term interlanguage development, and emotional responses to writing. Ultimately, the research seeks to contribute to theories of AI-mediated second language acquisition and offer practical recommendations for integrating AI and human feedback effectively in higher education L2 writing instruction.

Keywords: automated written corrective feedback, second language writing, learner autonomy, writing anxiety, EFL, longitudinal research, higher education.

الملخص

تستكشف هذه الدراسة الطولية متعددة المناهج التأثير طويل المدى لأدوات التغذية الراجعة المولدة بالذكاء الاصطناعي، مثل ChatGPT وGrammarly، على تطور الكتابة الأكاديمية لدى طلبة اللغة الإنجليزية كلغة أجنبية أو ثانية في المرحلة الجامعية على مدى ثلاث سنوات أكاديمية. وتهدف الدراسة إلى معالجة الفجوات الموجودة في أبحاث التغذية الراجعة التصحيحية الآلية للكتابة، من خلال التركيز على ثلاثة أبعاد رئيسية: التطور اللغوي، واستقلالية المتعلم، والنتائج الوجدانية. وسيتم تتبع نحو 80 طالبًا جامعيًا عبر ستة فصول دراسية باستخدام أدوات كمية لقياس التعقيد التركيبي، والتنوع المعجمي، ومعدلات الأخطاء، وقلق الكتابة، والكفاءة الذاتية، والاستمتاع بتعلم اللغة الأجنبية، إلى جانب بيانات نوعية مستمدة من المقابلات، واليوميات التأملية، وجلسات الاستدعاء المحفّز، وسجلات تفاعل الشاشة. وتسعى الدراسة إلى تقديم فهم أعمق لكيفية تفاعل المتعلمين مع التغذية الراجعة المولدة بالذكاء الاصطناعي وتأثيرها في ممارسات المراجعة، وتطور اللغة البيئية على المدى الطويل، والاستجابات العاطفية المرتبطة بالكتابة. وفي النهاية، تهدف الدراسة إلى الإسهام في تطوير نظريات

اكتساب اللغة الثانية المدعوم بالذكاء الاصطناعي، وتقديم توصيات عملية لدمج التغذية الراجعة البشرية والذكاء الاصطناعي بصورة فعالة في تدريس الكتابة باللغة الثانية في مؤسسات التعليم العالي.

الكلمات المفتاحية: الملاحظات التصحيحية الكتابية الآلية، كتابة اللغة الثانية، استقلالية المتعلم، قلق الكتابة، اللغة الإنجليزية كلغة أجنبية، بحث تنبئي، التعليم العالي.

1. Introduction

The emergence of generative artificial intelligence (AI) tools has introduced new possibilities and challenges for second language (L2) writing instruction [1]. Large language model (LLM)-based applications such as ChatGPT, Grammarly, and QuillBot provide learners with immediate, individualized, and low-anxiety feedback environments that extend writing support beyond traditional classroom settings [2][3]. In higher education contexts, where learners are expected to develop academic writing proficiency rapidly, these tools have become increasingly attractive as supplementary or alternative feedback mechanisms.

Recent studies suggest that AI-generated feedback can effectively improve surface-level writing features, including grammatical accuracy, lexical diversity, and sentence structure [4][5]. Automated written corrective feedback (AWCF) systems have evolved considerably from earlier rule-based technologies, now offering context-sensitive and conversational feedback capable of addressing more nuanced aspects of learner writing. However, current empirical evidence remains heavily concentrated on short-term interventions and cross-sectional designs, limiting understanding of how AI-mediated feedback influences writing development over extended periods.

This limitation is particularly significant because second language writing development is inherently longitudinal [6]. Improvements observed after brief AI interventions may reflect temporary performance gains rather than durable interlanguage development. Moreover, existing research has largely prioritized measurable linguistic outcomes while overlooking equally important learner-centered dimensions, such as autonomy, self-regulation, and emotional response to feedback.

Learner autonomy has emerged as a central concern in AI-assisted language education. On one hand, AI tools may promote independent revision behavior by enabling learners to seek feedback on demand, experiment with multiple drafts, and engage in self-directed problem-solving. On the other hand, excessive reliance on AI-generated suggestions may encourage passive dependence and reduce critical engagement with language choices. Understanding how learners negotiate this balance over time is essential for evaluating the pedagogical value of AI-mediated feedback.

In addition to autonomy, affective factors such as writing anxiety, self-efficacy, and foreign language enjoyment may significantly shape learners' interactions with AI tools. Research indicates that AI-generated feedback can reduce anxiety by providing private, immediate, and non-judgmental responses, thereby lowering the social pressure often associated with teacher or peer evaluation. Yet little is known about whether these affective benefits persist longitudinally or how they interact with writing development and feedback engagement.

To address these gaps, the present study proposes a longitudinal mixed-methods investigation of undergraduate EFL/ESL learners using AI-generated feedback across six consecutive semesters. Drawing on frameworks from automated written corrective feedback research, learner autonomy theory, and affective models of second language acquisition, the study examines how AI-mediated feedback shapes linguistic development, autonomous writing behavior, and emotional trajectories over time.

This research seeks to answer the following questions:

1. How do syntactic complexity, lexical diversity, and error frequency in undergraduate EFL learners' academic writing change across six semesters of AI-assisted feedback instruction?
2. Is there a significant difference in writing development trajectories between learners receiving AI-only feedback and those receiving hybrid (AI plus human tutor) feedback?
3. How do learners experience and describe the development of autonomy when engaging with AI-generated feedback over time?
4. What is the longitudinal affective trajectory of learners receiving AI-generated feedback, and how do affective variables interact with writing performance outcomes?
5. What individual, contextual, and technological factors mediate the relationship between AI feedback engagement and long-term writing development?

By addressing these questions, this study aims to contribute to emerging scholarship on AI in second language education and provide evidence-based recommendations for integrating generative AI responsibly into higher education writing pedagogy.

2. Material and methods

2.1 Research Design

This study adopts a longitudinal concurrent mixed-methods design to investigate the long-term effects of AI-generated feedback on second language (L2) writing development, learner autonomy, and affective response

among undergraduate EFL/ESL learners [7]. The study spans six consecutive academic semesters over a three-year period, allowing for the examination of developmental trajectories across time rather than isolated instructional outcomes.

The mixed-methods approach is grounded in a pragmatic research paradigm, which recognizes the value of integrating quantitative and qualitative methods to capture the complexity of language learning processes. Quantitative measures will provide evidence of longitudinal changes in writing performance and affective variables, while qualitative inquiry will offer insight into learners' experiences, perceptions, and evolving engagement with AI-mediated feedback.

The concurrent design enables both data strands to be collected and analyzed in parallel, facilitating continuous triangulation throughout the study. This design is particularly suitable for L2 writing research because writing development involves dynamic interactions among linguistic, cognitive, and emotional dimensions that cannot be fully understood through a single methodological lens.

2.2 Participants and Context

Participants will be recruited from undergraduate academic writing courses at a research-intensive university operating within a multilingual higher education environment. The target sample consists of approximately 80 first-year EFL/ESL learners enrolled in English-medium academic writing programs who will be followed longitudinally through the duration of their undergraduate studies.

To ensure diversity within the sample, participants will be stratified according to:

- first language background,
- English proficiency level at program entry,
- and prior experience with AI writing technologies.

Initial English proficiency will be determined through institutional placement measures equivalent to IELTS band scores ranging from approximately 5.5 to 7.5. Participants will engage regularly with at least one AI-generated feedback tool, including ChatGPT and/or Grammarly, as part of their writing instruction and revision practices.

Two instructional conditions will be examined:

1. AI-only feedback condition, in which learners primarily rely on AI-generated corrective feedback.
2. A hybrid feedback condition, combining AI-generated feedback with teacher-mediated guidance and commentary.

This comparison will allow the study to evaluate whether hybrid pedagogical approaches produce stronger long-term outcomes than AI-supported instruction alone.

2.3 Instruments and Data Collection

Data collection will occur across five interconnected streams designed to capture linguistic development, learner behavior, and affective change over time.

Writing Corpus

The primary quantitative dataset will consist of a longitudinal corpus of student writing collected across six semesters. The corpus is expected to include approximately 1,400–1,500 academic texts produced through regular coursework assignments. These texts will be analyzed for:

- syntactic complexity,
- lexical diversity,
- cohesion,
- error frequency.

Automated text analysis tools such as Coh-Metrix and TAALES will be used to generate quantitative linguistic measures, while manual error tagging procedures based on established L2 error taxonomies will be employed to examine grammatical accuracy.

Affective Surveys

To examine emotional dimensions of AI-mediated writing development, standardized affective instruments will be administered at multiple time points during the study. These instruments will assess:

- second language writing anxiety,
- writing self-efficacy,
- foreign language enjoyment.

The survey protocol will include validated scales widely used in second language acquisition research, including the Second Language Writing Anxiety Inventory (SLWAI), the Writing Self-Efficacy Scale, and the Foreign Language Enjoyment Scale. Repeated administration across semesters will enable the tracking of affective trajectories over time.

Screen-Capture Interaction Logs.

A stratified sub-sample of approximately 20 participants will participate in screen-capture recording sessions during writing and revision tasks. Recording software will document real-time interactions with AI feedback systems, including:

- revision decisions,
- feedback consultation patterns,
- time-on-task,
- interaction sequences.

These recordings will provide insight into how learners operationalize AI-generated feedback during authentic writing processes.

Semi-Structured Interviews

Semi-structured interviews will be conducted with a purposively selected group of participants representing variation in proficiency level, linguistic background, and affective profile. Interviews will occur at the end of each academic year, generating multiple cycles of longitudinal qualitative data.

Interview protocols will focus on:

- learner perceptions of AI feedback,
- evolving experiences of autonomy,
- emotional responses to AI-assisted writing,
- perceived changes in writing practices over time.

Stimulated Recall Sessions

Stimulated recall protocols will be conducted following selected writing sessions. Participants will review segments of their own screen-capture recordings and verbally reflect on the cognitive and emotional processes underlying their interactions with AI-generated feedback.

This method is intended to reveal otherwise inaccessible aspects of learner decision-making, including:

- reasons for accepting or rejecting AI suggestions,
- moments of uncertainty or confidence,
- strategies for evaluating feedback quality.

2.4 Data Analysis

Quantitative and qualitative data will be analyzed separately before integration during the interpretation stage.

Quantitative Analysis

Longitudinal quantitative data will be analyzed using growth curve modeling (GCM) within a multilevel regression framework to examine developmental trajectories in writing performance across time. This approach will allow the study to capture both group-level trends and individual variation in writing development.

Repeated-measures analyses of variance (ANOVA) and multivariate analyses (MANOVA) will be employed to compare outcomes across instructional conditions and time points. Structural equation modeling (SEM) will also be used to investigate relationships among affective variables, learner autonomy, and writing performance.

Qualitative Analysis

Qualitative data derived from interviews, stimulated recall sessions, and reflective accounts will be analyzed using reflexive thematic analysis. The analysis will focus on identifying recurring themes related to:

- learner autonomy,
- AI dependence,
- revision behavior,
- affective response,
- perceptions of feedback usefulness.

Interaction analysis will additionally be applied to screen-capture data to examine patterns of engagement with AI-generated feedback during writing tasks.

Integration of Findings

Integration of quantitative and qualitative findings will occur through joint data displays and comparative interpretation. Qualitative findings will be used to contextualize and explain quantitative patterns, particularly in cases where developmental trajectories differ substantially across participants or instructional conditions.

This integrative approach is intended to provide a comprehensive understanding of how linguistic, behavioral, and affective dimensions of AI-mediated writing development interact over time.

2.5 Ethical Considerations

Ethical approval will be obtained from the university's institutional research ethics board prior to data collection. Participation in the study will be voluntary, and all participants will provide informed written consent.

To protect participant confidentiality:

- pseudonyms will replace identifying information in all datasets and publications,

- digital data will be securely encrypted and stored,
- and participants will retain the right to withdraw from the study at any stage without academic penalty.

Given the longitudinal nature of the research and the collection of screen-capture data, particular attention will be paid to issues of privacy, informed consent, and responsible data management. The researcher will also maintain a reflexive stance throughout the project to acknowledge potential positionality and interpretive bias.

Results and discussion

The study is expected to demonstrate that sustained engagement with AI-generated feedback contributes positively to multiple dimensions of second language (L2) writing development over time. Specifically, learners using AI-assisted feedback are anticipated to show measurable improvement in syntactic complexity, lexical diversity, and grammatical accuracy across six semesters of academic writing instruction [4][5].

However, these gains are not expected to occur uniformly across all dimensions of writing. Improvements are likely to be strongest in lower-order linguistic features, such as grammar correction, sentence structure, and vocabulary use, while more gradual or inconsistent development may emerge in higher-order concerns including rhetorical organization, coherence, and argumentative sophistication. This distinction may reinforce existing claims that AI-generated feedback is particularly effective for micro-level revision but less capable of supporting advanced academic discourse development independently.

Differences are also expected between instructional conditions. Learners participating in hybrid feedback environments combining AI-generated and human-mediated feedback are anticipated to demonstrate stronger long-term writing outcomes than those relying primarily on AI feedback alone. Hybrid models may provide complementary advantages by allowing AI systems to address surface-level corrections while instructors focus on discourse-level guidance, critical thinking, and disciplinary writing conventions.

Regarding learner autonomy, the study is expected to reveal increasingly sophisticated self-regulated writing behaviors among participants over time. Learners may demonstrate greater capacity for independent revision, strategic feedback evaluation, and iterative drafting practices as familiarity with AI tools increases. At the same time, findings may also indicate that excessive dependence on AI-generated suggestions can inhibit critical engagement and reduce learner agency in certain contexts, particularly among lower-proficiency participants.

In terms of affective outcomes, AI-generated feedback is expected to reduce writing anxiety and increase writing self-efficacy and foreign language enjoyment across the duration of the study. The immediate, individualized, and non-judgmental nature of AI feedback may create a psychologically safer revision environment than traditional teacher-centered correction practices. Nevertheless, the study may also identify affective variability linked to learners' technological confidence, feedback literacy, and prior experience with AI systems.

Qualitative findings are expected to reveal substantial variation in how learners interpret and utilize AI-generated feedback. Some participants may engage critically and strategically with AI suggestions, while others may adopt more passive revision behaviors characterized by uncritical acceptance of automated corrections. Such differences are likely to be influenced by individual learner characteristics, instructional scaffolding, and contextual factors within the learning environment.

Overall, the anticipated findings are expected to support the argument that AI-generated feedback can facilitate meaningful long-term writing development when integrated thoughtfully within pedagogically guided learning environments [8]. However, the study is also expected to demonstrate that AI systems alone are insufficient substitutes for human instruction and that successful implementation depends heavily on learner training, feedback literacy, and balanced pedagogical design.

Conclusion

This study seeks to contribute to the growing body of scholarship examining the role of artificial intelligence in second language writing instruction by providing longitudinal evidence regarding the developmental, affective, and autonomous dimensions of AI-mediated feedback. In contrast to much of the existing literature, which relies heavily on short-term intervention studies, the present research emphasizes sustained writing development across an extended academic timeframe.

One anticipated contribution of the study lies in its potential to clarify whether improvements associated with AI-generated feedback represent temporary performance effects or genuine interlanguage development. If learners demonstrate stable longitudinal gains in writing accuracy and complexity, the findings may support arguments that AI-mediated feedback can contribute meaningfully to durable language acquisition processes rather than merely assisting short-term task completion.

The study may also extend theoretical understandings of learner autonomy in AI-supported educational environments. Existing debates frequently position AI either as a tool that enhances self-regulated learning or as a mechanism that promotes learner dependency. The expected findings suggest that this relationship is more nuanced and dynamic. AI-generated feedback may foster autonomy when learners engage critically with feedback and develop revision strategies, but it may undermine autonomy when learners become overly reliant on automated correction without reflective engagement.

A further contribution concerns the affective dimensions of L2 writing development. The anticipated reduction in writing anxiety and increase in self-efficacy may reinforce emerging research suggesting that AI-generated feedback creates psychologically supportive writing environments. Such findings could have important implications for understanding how emotional variables mediate learner engagement with revision processes and influence long-term writing development [9].

Pedagogically, the study is expected to support the implementation of hybrid feedback models integrating AI-generated corrective feedback with teacher-mediated instruction. Rather than replacing instructors, AI tools may function most effectively when embedded within balanced pedagogical frameworks that combine technological efficiency with human expertise, mentorship, and rhetorical guidance. These findings could help institutions design more sustainable and pedagogically responsible approaches to AI integration in academic writing curricula. The research may additionally contribute methodological value by demonstrating the usefulness of longitudinal mixed-methods designs in investigating AI-assisted language learning. By integrating linguistic analysis, affective measurement, and qualitative inquiry, the study may offer a more comprehensive account of AI-mediated writing development than approaches relying exclusively on quantitative performance measures.

Despite its anticipated contributions, the study will inevitably face several limitations. Participant attrition over the three-year research period may affect sample consistency, while rapid technological developments in generative AI may alter the capabilities of feedback systems during the course of the study itself. Furthermore, contextual variables such as instructor practices, institutional policies, and learner access to technology may influence outcomes in ways that are difficult to fully control.

Future research may therefore expand upon this study by exploring AI-mediated writing development across different educational contexts, proficiency levels, and linguistic backgrounds. Additional investigation into discipline-specific writing practices, multilingual learners, and emerging multimodal AI systems may further deepen understanding of how artificial intelligence reshapes second language writing pedagogy.

As generative AI continues to transform higher education, the findings of this study are expected to provide timely evidence regarding both the opportunities and limitations of AI-generated feedback in second language writing instruction. Ultimately, the study argues for a balanced and critically informed approach to AI integration—one that prioritizes learner development, pedagogical intentionality, and the continued importance of human guidance within academic writing education [8].

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, 100745.
- [2] Guo, Y., & Wang, X. (2024). Integrating AI feedback in L2 writing: Peer review and hybrid feedback models. *TESOL Quarterly*, 58(1), 112–139.
- [3] Mynard, J. (2019). *Promoting learner autonomy and self-directed learning*. Palgrave Macmillan.
- [4] Wu, H., Garza, E. V., & Guzman, N. (2023). Comparing the effectiveness of ChatGPT and Grammarly as automated writing evaluation tools in an EFL context. *Education Sciences*, 13(4), 388.
- [5] Steiss, J., Tate, T., Graham, S., et al. (2024). Comparing the quality of human and ChatGPT feedback on students' writing. *Learning and Instruction*, 91, 101894.
- [6] Cambridge Core. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 189–208.
- [7] Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- [8] Han, Y., & Sari, E. (2024). Combined AWE and teacher feedback in L2 writing courses. *Language Learning & Technology*, 28(1), 45–67.
- [9] Ferris, D. R. (2004). The grammar correction debate in L2 writing: Where are we, and where do we go from here? *Journal of Second Language Writing*, 13(1), 49–62.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **AJASHSS** and/or the editor(s). **AJASHSS** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.