Available Online: 14 October 2026 DOI: 10.54254/2753-7080/2026.37885

The effects of artificial intelligence feedback on the writing proficiency of English majors in a Chinese context

***Jie Li*[1,*], *Supei Wang*[2]**

[1]Foreign Languages College, Inner Mongolia University, Hohhot, China
[2]Huayan Middle School, Erdos, China

*Corresponding Author. Email: lijane1314@163.com

Abstract. Artificial intelligence feedback has become a hot topic in second language writing research. Although previous studies have yielded fruitful findings, most of them have adopted a cross-sectional research design, failing to demonstrate the long-term effects of AI feedback. In view of this, the current study adopts a longitudinal design to study the effects of AI feedback on the writing proficiency of 72 English majors at a comprehensive university in northwest China. The writing proficiency was evaluated via lexical complexity, syntactic complexity and accuracy. Findings revealed that AI feedback had a significant positive effect on English majors' language accuracy and lexical complexity, but had no effect on their syntactic complexity. This study finally highlights a few implications for the use of AI in second language writing.

Keywords: AI feedback, second language writing, lexical complexity, syntactic complexity, accuracy

1. Introduction

In Second Language (L2) writing, feedback is of great importance, as it helps learners identify areas for improvement [1]. Traditionally, feedback is given by teachers, yet they often face constraints such as large class sizes and slow responses [2, 3]. These constraints have prompted researchers to seek alternative feedback provision methods. With the rapid development of Artificial Intelligence (AI), AI tools have become a viable option, which can provide real-time and detailed feedback [4]. Given this potential, researchers worldwide have conducted empirical studies to investigate the effects of AI feedback on learners' writing performance [5-7]. Though these previous studies have achieved fruitful findings, confirming the positive effects of AI feedback on learners' writing performance, most of them have adopted a cross-sectional research design, with little attention paid to its long-term effects on L2/EFL learners' writing proficiency. Additionally, most previous studies focused on the overall quality of learners' writing, and did not specifically examine the changes across different linguistic dimensions, such as complexity, accuracy, and fluency, which are widely used indices in second language writing research [8]. In order to fill this lacuna, this study adopts a longitudinal design to investigate the effects of AI feedback on Chinese English majors' writing proficiency indexed by complexity and accuracy. This study excludes fluency in that previous research has employed inconsistent indicators, making cross-study comparisons impossible [9].

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

2. Literature review

2.1. Teacher feedback in L2 writing

In the field of L2 writing instruction, feedback, as an important intervention to improve learners' writing skills, has received widespread research attention in applied linguistics and second/foreign language education. Previous research has explored L2 writing feedback from multiple perspectives, which can be broadly categorized into four dominant research strands.

First, numerous studies have investigated the fundamental effects of feedback, including feedback effectiveness [10], comparison of feedback strategies [11], feedback timing [12], factors influencing feedback effectiveness [13], and dynamic feedback [14]. Second, a body of research has focused on stakeholder perceptions, covering learners' beliefs and attitudes towards feedback [15], and learners' engagement with feedback [16]. Third, several studies have analyzed the linguistic features of feedback, particularly the implicit attitudes conveyed through feedback discourse [17]. Fourth, emerging research has centered on the feedback literacy, examining the literacy development of both teachers and learners in feedback practices [18].

Overall, effective feedback helps monitor learners' writing process, deepens learners' understanding of writing norms, and improves the accuracy, complexity, and overall quality of learners' writing. However, scholars both domestically and internationally have not yet reached a consensus on how to provide feedback, when to provide feedback, and which feedback strategies are effective, so further research is needed [1].

2.2. Automatic feedback in L2 writing

With the advancement of computer-assisted language learning and teaching, online automatic corrective feedback delivered by Automated Writing Evaluation (AWE) systems has evolved within this field. Nevertheless, only a handful of scholars have conducted preliminary investigations into this topic to date [19, 20]. Although AWE holds substantial potential for providing feedback on L2 writing, some researchers have raised doubts regarding its pedagogical effectiveness [21].

In recent years, breakthrough developments in AI technologies have opened new avenues for L2 writing feedback. Large language models built on natural language processing techniques, such as ChatGPT, have attracted extensive scholarly attention owing to their capacity to deliver efficient feedback, giving rise to a growing body of relevant research. These studies mainly centered on the comparison of ChatGPT-generated feedback with teacher feedback [5, 22-24], the effectiveness of ChatGPT-generated feedback [25-27], and learners' attitudes and affective experiences toward ChatGPT-generated feedback [5, 28, 29]. For example, Alsofyani and Barzanji [5] adopted an explanatory sequential mixed-methods design to compare the effects of ChatGPT-generated feedback and conventional teacher feedback among 102 intermediate-level Saudi female tertiary EFL learners over an eight-week intervention. The 102 participants were randomly assigned to an experimental group receiving ChatGPT-generated feedback, and a control group receiving teacher feedback. The quantitative results revealed no statistically significant post-test differences in overall writing performance between the two groups, suggesting that ChatGPT-generated feedback was as effective as teacher feedback. Unlike Alsofyani and Barzanji [5] who focusing on assessing the improvements of the participants' writing performance, Algaraady and Mahyoob [22] examined the quality of feedback generated by ChatGPT and instructors. They found that ChatGPT successfully detected such surface level errors in the participants' written texts as punctuation, spelling, vocabulary, grammar, and sentence structure, yet failed to identify deep structural and pragmatic issues. This finding demonstrates that ChatGPT cannot replace instructors when addressing much deeper dimensions of learners' writing. Even so, Meyer et al. [30] found that generative AI feedback could foster students' writing motivation and positive emotions.

Overall, existing research has preliminarily validated the positive effects of AI tools such as ChatGPT in L2 writing feedback. Still, a couple of limitations remain. For one thing, previous studies overwhelmingly focused on ChatGPT, with markedly insufficient attention paid to other large language models, such as DeepSeek, GLM-4, Kimi, Doubao, and ERNIE Bot. Empirical studies are still needed to examine these models' performance in real-time corrective feedback, the efficacy of their feedback strategies, and their capacity to address learners' writing problems. For another, most available studies adopted relatively short research cycles (typically less than eight weeks). This has made it difficult to fully capture the long-term effects of AI feedback on learners' writing proficiency. In view of these two lacunas in existing studies, the current study adopts an academic-year-long research design to investigate the long-term effects of ERNIE Bot (Wen Xin Yi Yan in Chinese), a representative large language model developed by Baidu in China, on English majors' writing proficiency in a Chinese context.

3. Methodology

3.1. Research questions

The current study answers the following two questions:

(1) How does AI feedback affect English majors' writing proficiency indexed by complexity over an academic year?

(2) How does AI feedback affect English majors' writing proficiency indexed by accuracy over an academic year?

3.2. Participants

This study was conducted in a foreign languages college at a comprehensive university in northwest China. It involved 72 English majors (62 females and 10 males) from three intact classes. They were between the ages of 19 and 22, with a history of 9 to 15 years of learning English and a similar learning background. At the time of the study, the participants were in their third year of a four-year undergraduate program. In this program, they were required to take compulsory and elective courses. Most of the required or recommended curricular resources are aimed at improving their English language proficiency. At the beginning of the study, they received an oral explanation of the research purpose as well as a promise to keep their identities confidential, and they all agreed to participate. In addition, none of the participants had used AI tools to get feedback for their writing texts.

3.3. Data collection

At the time of the data collection, the participants took two courses where the first researcher was their instructor, aiming to help improve their English teaching skills and academic writing skills. The participants were required to write an argumentative composition on an assigned topic as an after-class assignment. Each composition followed the same process: The participants wrote their first drafts after the instructor assigned the topic, then uploaded them to ERNIE Bot to obtain feedback. Based on AI feedback, the participants corrected their compositions. Finally, they submitted their first and second drafts as well as the records of their interaction process with AI.

The data collection period lasted for a whole academic year, approximately 36 weeks. Each participant wrote one argumentative composition every three weeks, yielding a total of 12 compositions over the

academic year. Ultimately, 864 compositions from 72 participants were collected, constituting the longitudinal corpus.

3.4. Data coding

To answer the research questions, the written texts completed by the participants were analyzed using quantitative measures which included measures of complexity and accuracy. Both complexity and accuracy measures were based on a count of T-units and clauses. A T-unit is defined as "one main clause plus whatever subordinate clauses happen to be attached to or embedded within it" [31]. A clause distinguishes between independent and dependent clauses. An independent clause is one which includes a finite verb and can be used on its own, while a dependent clause must be used with another clause to form a grammatical sentence, which, following Foster et al. [32], is defined as one containing a finite or non-finite verb plus one such additional clause element as subject, object, complement, or adverbial.

Following previous research [8], complexity included syntactic complexity and lexical complexity. Syntactic complexity was measured via two indices. One was the proportion of Words to T-units (W/T), which was calculated by dividing total words by total T-units. The other one was the proportion of Clauses to T-units (C/T), which was calculated by dividing total clauses by total T-units. These two indices have been established as reliable ones in L2 writing research [33], which examine different subconstructs of syntactic complexity [34]. Lexical complexity was measured using the Corrected Type Token Ratio (CTTR), which, unlike simple TTR value, could avoid the influence of text length [35, 36]. To measure accuracy, two indices were used: the proportion of Error-Free T-units to all T-units (EFT/T) and the proportion of Error-Free Clauses to all Clauses (EFC/C). Errors in this study, as Storch [37] did in her research, included the ones in lexis, syntax and morphology, such as word choice, word order, verb tense, subject-verb agreement, missing elements, to name just a few. The errors in spelling and punctuation were excluded.

3.5. Analysis

To evaluate the changes in the participants' writing proficiency, the researchers, following the analytical procedure adopted in previous research [38], computed the difference between starting and final writing proficiency. The starting writing proficiency was quantified as the mean value of the first two writing texts, and the final writing proficiency was quantified as the mean value of the last two writing texts. In order to compare the differences in mean values, paired-samples t-test was performed via SPSS 27. The alpha level was set at 0.05. Cohen's *d* was computed to obtain effect size estimates, which allowed for interpreting the magnitude of the observed differences [39]. Following Plonsky and Oswald [40], within-group effect sizes were interpreted as small at $d = 0.6$, medium at $d = 1.0$, and large at $d = 1.4$.

4. Results

4.1. Effects of AI feedback on complexity

Complexity included lexical complexity and syntactic complexity, which were measured by CTTR, W/T and C/T respectively. The following three tables present their descriptive and inferential statistics.

Table 1. Descriptive and inferential statistics for CTTR

Time	*M*	*SD*	*MD*	*t* (71)	*d*
Starting	5.309	0.416	-0.606	-9.847*	1.375
Final	5.915	0.464			

Note: * indicates statistical significance at the 5% level.

Table 2. Descriptive and inferential statistics for W/T

Time	*M*	*SD*	*MD*	*t* (71)	*d*
Starting	15.694	2.429	0.248	0.710	0.104
Final	15.446	2.333			

Table 3. Descriptive and inferential statistics for C/T

Time	*M*	*SD*	*MD*	*t* (71)	*d*
Starting	2.279	0.380	-0.032	-0.545	0.085
Final	2.311	0.374			

Table 1 shows that CTTR scores increased from the starting to the final measurement. The mean value rose by 0.606 points, from 5.309 to 5.915. This difference was statistically significant ($p < 0.05$), with a large effect size ($d = 1.375$). This suggests that AI feedback produced a significant positive effect on the lexical complexity of the learners' writing.

W/T values, however, did not change substantially across the academic year. As shown in Table 2, the means were 15.694 and 15.446 at the initial and final measurement respectively, indicating a non-significant difference ($t(71) = 0.710$, $p > 0.05$) and a negligible effect size ($d = 0.104$). Although AI feedback enhanced learners' vocabulary use, it exerted no significant influence on syntactic complexity as measured by words per T-unit. Table 3 presents the C/T results, which similarly revealed no significant change. From the initial to the final measurement, the ratio remained stable ($t(71) = -0.545$, $p > 0.05$, $d = 0.085$). Collectively, these two indices suggest that AI feedback exerted no effect on the learners' syntactic complexity.

4.2. Effects of AI feedback on accuracy

Regarding accuracy, EFT/T and EFC/C ratios were calculated to assess whether students produced fewer errors in their T-units and clauses after receiving AI feedback over an academic year. Their descriptive and inferential statistics are shown in Table 4 and Table 5.

Table 4. Descriptive and inferential statistics for EFT/T

Time	*M*	*SD*	*MD*	*t* (71)	*d*
Starting	0.717	0.207	-0.105	-3.759*	0.492
Final	0.822	0.220			

Note: * indicates statistical significance at the 5% level.

Table 5. Descriptive and inferential statistics for EFC/C

Time	*M*	*SD*	*MD*	*t* (71)	*d*
Starting	0.861	0.120	-0.054	-3.063*	0.444
Final	0.915	0.123			

Note: * indicates statistical significance at the 5% level.

Accuracy measures presented in Table 4 and Table 5 revealed improvements from the initial to the final measurement. The EFT/T ratio increased from 0.717 to 0.822, an increase of 0.105 points. This gain was statistically significant ($t(71) = -3.759$, $p < 0.05$), with a small effect size ($d = 0.492$). This indicates that learners produced fewer errors in their T-units after receiving AI feedback. Additionally, as illustrated in Table 5, EFC/C scores increased from 0.861 to 0.915 over the academic year intervention. This improvement was statistically significant ($t(71) = -3.063$, $p < 0.05$), with a small effect size ($d = 0.444$). These results suggest that AI feedback significantly improved the accuracy of the learners' written outputs.

5. Discussion

The current study aimed to investigate the long-term effects of AI feedback on Chinese English majors' writing proficiency. With respect to the first research question concerning the long-term effects of AI feedback on the learners' writing complexity, the study found that AI feedback had a significant positive effect on the participants' lexical complexity indexed by CTTR. This indicates that the participants were able to draw upon a wide range of words instead of repeatedly relying on a limited number of words. This improvement may be due to the input the participants received through their interactions with AI. According to Krashen [41], greater amounts of comprehensible input obtained through reading are linked to the development of vocabulary knowledge, spelling ability, and other dimensions of linguistic competence. In the current study, the participants were required to revise their initial drafts based on AI-generated feedback. In order to complete this task, they needed to understand the AI feedback they received, which may function as one source of comprehensible input. Thus, they were motivated to process such input, or even incorporate suggestions from the feedback into their revised texts, which, in turn, may have facilitated their lexical growth in writing.

In addressing the first research question, the study also found that AI feedback had no measurable effect on the syntactic complexity of the participants' written outputs. A major factor that needs to be mentioned here is that the current study, unlike previous studies [5, 26], did not organize planned training for the participants before the intervention. Thus, the participants may have trouble in providing appropriate prompts to help them make full use of AI's potential to generate high-quality feedback. Moreover, AI, despite its benefits, also has limitations [27]. Taken together,When AI generated feedback, its corrections might largely center on word-level choices and grammatical errors, offering relatively limited guidance for expanding sentence structures or embedding more subordinate clauses. As a result, upon receiving AI feedback, the participants tended to correct identified errors rather than restructure their sentences. This implies that teachers should assist students to develop their AI literacy and competence to utilize AI tools effectively.

With regards to the second research question concerning the long-term effects of AI feedback on the learners' writing accuracy, the current study found that AI feedback significantly improved the participants' writing accuracy measured by EFT/T and EFC/C. This finding is consistent with previous research [25]. Boudouaia et al. [25] conducted an experiment with 76 undergraduates from a private university in Algeria to investigate the effects of ChatGPT-4 on students' EFL writing. They found that ChatGPT-4 enhanced students'

proficiency in using a variety of grammatical structures accurately. This finding is also similar to Masoudi's study [27], which revealed that ChatGPT could identify and correct spelling and grammatical mistakes the students made in their written texts, and suggest alternative words and phrases to enhance vocabulary as well as improving sentence structure, coherence and clarity in writing. The significant positive changes in the participants' accuracy could be attributable to a couple of factors. First, AI feedback, unlike traditional teacher feedback, is immediate rather than delayed. This has enabled the participants to detect and rectify the mistakes they made promptly, which, in turn, might help to prevent misconceptions from becoming ingrained and facilitate continued progress [6]. Furthermore, AI feedback can be tailored to individual learners. This degree of personalization is difficult to realize through conventional feedback practices. By addressing each learner's strengths and weaknesses, AI can generate targeted support within their zone of proximal development [5], hence assisting them in advancing their writing accuracy. Finally, AI-generated feedback, more often than that, supplies metalinguistic explanations which clarify the rules underlying grammatical mistakes [26]. Such explanatory feedback could help the participants internalize linguistic rules, which might transfer to new writing assignments and improve their long-term writing accuracy.

6. Conclusion

The aim of the current study was to examine the long-term effects of AI feedback on English majors' writing proficiency indexed by complexity and accuracy in a Chinese context. Findings revealed that AI feedback significantly enhanced the participants' lexical complexity and writing accuracy. Nevertheless, it exerted no statistically measurable effects on the syntactic complexity of the participants' written outputs over the academic year. This suggests that AI, to a larger or lesser degree, is effective in providing feedback in EFL writing teaching, so EFL teachers and learners can use it in their teaching and learning. Even so, practitioners should not overlook its potential limitations. Therefore, it is essential for both EFL teachers and learners to receive training to improve their AI literacy. By strengthening such literacy, they are able to use AI tools critically, reflectively, and effectively, which, finally helps promote EFL learners' comprehensive development in L2 writing.

The study is subject to two limitations. First, convenience sampling was adopted for participant recruitment. The current study was conducted at a comprehensive university in China, with a sample consisting of 72 English major students. This may limit the applicability of the research findings to non-English majors or English majors from diverse geographical regions and cultural backgrounds. The other limitation of the current study is that it did not delve into the dynamic developmental process of the participants' writing proficiency.

Given the limitations mentioned above, future studies could adopt random sampling and recruit participants from multiple universities across different regions, or with varying proficiency levels. This would help validate whether the observed effects of AI feedback can be generalized to a broader range of EFL learners. Future investigations could also employ longitudinal case studies or mixed-methods design to trace how learners' writing proficiency evolves over the course of AI-feedback interventions. Capturing the dynamic developmental trajectory would provide deeper insights into how AI-generated feedback shapes L2 writing development.

Funding

The Foreign Languages Special Project of the Social Science Fund of Inner Mongolia Autonomous Region (2026MY04)

References

[1] Miao, J., Chang, J., & Ma, L. (2023). Research trends of written corrective feedback in L2 writing: A bibliometric analysis. *SAGE Open, 13*(1), 1-13.

[2] Li, S. (2025). Generative AI and second language writing. *Digital Studies in Language and Literature*, 1-30.

[3] Yao, Y., Zhu, X., Xiao, L., & Lu, Q. (2025). Secondary school English teachers' application of artificial intelligence-guided chatbot in the provision of feedback on student writing: An activity theory perspective. *Journal of Second Language Writing, 67*, 1-18.

[4] Karagoz, I. (2025). AI-generated feedback in English writing instruction for language learners: A systematic review. *The Reading Matrix, 25*(1), 51-67.

[5] Alsofyani, A. H., & Barzanji, A. M. (2025). The effects of ChatGPT-generated feedback on Saudi EFL learners' writing skills and perception at the tertiary level: A mixed-methods study. *Journal of Educational Computing Research, 63*(2), 431-463.

[6] Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia, 15*(7), 1-26.

[7] Sun, P. P., Xu, J., & Zhang, L. J. (2025). Exploring the value-added benefits of GAI feedback on English writing quality among Chinese university students. *Foreign Language Education in China, 8*(4), 24-36.

[8] Fernández Dobao, A. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work. *Journal of Second Language Writing, 21*, 40-58.

[9] Zhang, M. X., & Plonsky, L. (2020). Collaborative writing in face-to-face settings: A substantive and methodological review. *Journal of Second Language Writing, 49*, 1-18.

[10] Lee, I. (2020). Utility of focused/comprehensive written corrective feedback research for authentic L2 writing classrooms. *Journal of Second Language Writing, 49*, 1-7.

[11] Sarre´, C., Grosbois, M., & Brudermann, C. (2021). Fostering accuracy in L2 writing: Impact of different types of corrective feedback in an experimental blended learning EFL course. *Computer Assisted Language Learning, 34*, 707–729.

[12] Shintani, N. (2016). The effects of computer-mediated synchronous and asynchronous direct corrective feedback on writing: A case study. *Computer Assisted Language Learning, 29*(3), 517–538.

[13] Zheng, Y., & Yu, S. (2018). Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assessing Writing, 37*, 13–24.

[14] Kurzer, K. (2018). Dynamic written corrective feedback in developmental multilingual writing classes. *TESOL Quarterly, 52*(1), 5-33.

[15] Zhang, Z. (2020). Engaging with automated writing evaluation (AWE) feedback on L2 writing: Student perceptions and revisions. *Assessing Writing, 43*, 78–91.

[16] Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing, 51*, 1-16.

[17] Mahfoodh, O. H. A. (2017). "I feel disappointed": EFL university students' emotional responses towards teacher written feedback. *Assessing Writing, 31*, 53-72.

[18] Lee, I., & Mao, Z. (2024). Writing teacher feedback literacy: Surveying second language teachers' knowledge, values, and abilities. *Journal of Second Language Writing, 63*, 1-7.

[19] Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing, 1*, 1–12.

[20] Barrot, J. S. (2021). Using automated written corrective feedback in the writing classrooms: Effects on L2 writing accuracy. *Computer Assisted Language Learning, 1*(1), 1–24.

[21] Strobl, C., Aihaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education, 131*, 33–48.

[22] Algaraady, J., & Mahyoob, M. (2023). ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journal (AWEJ) Special Issue on CALL*, 3–17.

[23] Guo, K., & Wang, D. L. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463.

[24] Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, 1–15.

[25] Boudouaia, A., Mouas, S., & Kouider, B. (2024). A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *Journal of Educational Computing Research, 62*(6), 1289–1317.

[26] Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments, 11*(9), 1-18.

[27] Masoudi, H. A. (2024). Effectiveness of ChatGPT in improving English writing proficiency among non-native English speakers. *International Journal of Educational Sciences and Arts, 3*(4), 62–84.

[28] Chan, C. K. Y., & Hu, W. J. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *Int J Educ Technol High Educ, 20*, 1-18.

[29] Xiao, Y. Y., & Zhi, Y. Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages, 8*, 1-12.

[30] 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*, 1-10.

[31] Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English, 43*(7), 732-739.

[32] Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375.

[33] Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing, 26*(3), 445-466.

[34] Norris, J., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578.

[35] Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42-65.

[36] Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language.* Oxford University Press.

[37] Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing, 14*, 153-173.

[38] Huang, T., Steinkrauss, R., & Verspoor, M. (2021). Variability as predictor in L2 writing proficiency. *Journal of Second Language Writing, 52*, 1-14.

[39] Xu, J. F., & Li, C. Y. (2022). Timing of form-focused instruction: Effects on EFL learners' grammar learning. *Studies in Second Language Learning and Teaching, 12*(3), 405-433.

[40] Plonsky, L., & Oswald, F. L. (2014). How big is "Big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878-912.

[41] Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal, 73*(4), 440-464.