Exploring the Role of AI Technology in Shaping College Students' English Writing Development: Insights from the Complex Dynamic Systems Theory

https://orcid.org/0000-0002-9640-6093

[1] Jian Wang
Geely University of China

https://orcid.org/0009-0002-0014-5093

[2] Xiaohong Zhang
Panzhihua University, China

Abstract

The proliferation of Artificial Intelligence (AI) technology has profoundly impacted English language teaching, especially within the realm of writing instruction. Despite extensive research exploring the application of AI in education, a significant gap remains in studies assessing the influence of AI-enhanced teaching on learners' writing proficiency. In this context, this study employed a longitudinal research design to examine the effects of integrating Kimi—an AI tool in China—into English writing instruction, framed within the Complex Dynamic Systems Theory. The participants, 12 tertiary-level English learners from the southwestern region of China, engaged with Kimi to complete weekly essay assignments over a 14-week period. They also participated in seven writing tests. The findings indicate that participants' overall writing proficiency, as well as lexical and syntactic accuracy and complexity of their test essays, exhibited an upward trajectory and significantly improved. However, insignificant improvement was observed in overall writing fluency. Additionally, improvement in learners' writing proficiency varied substantially among individual learners. This study provides compelling empirical evidence for English-as-a-foreign-language (EFL) writing pedagogy, highlighting the crucial roles of differentiated instruction, dynamic assessment, and technology-assisted tools in enhancing students' writing proficiency. English teachers are urged to adapt their pedagogical approaches flexibly, leveraging these insights to support the holistic development of each learner.

Keywords

AI Technology; Complex Dynamic Systems Theory; EFL Learners; Writing Development

Research background

English writing proficiency, as a productive skill, represents an essential dimension of overall English language competence (Bai & Hu, 2017). However, traditional English writing instruction confronts a myriad of challenges that impede effective learning and teaching. First and foremost, in conventional writing classrooms, the pedagogical process is predominantly teacher-centered, with instructors acting as

---

[1] Wang, J. School of Foreign Languages and Cultures, Geely University of China. Email: wangjian@guc.edu.cn. ORCID: https://orcid.org/0000-0002-9640-6093

[2] Zhang, X. School of Foreign Languages and Cultures, Panzhihua University, China. Email: zhangxiaohong2024@163.com. ORCID: https://orcid.org/0009-0002-0014-5093

the primary purveyors of language input and writing knowledge. Teachers often function as knowledge transmitters, rigidly dictating the content, pace, and methods of learning, while students remain largely passive recipients. This imbalance results in diminished student engagement, curtailed autonomy, and waning interest in writing. Moreover, the onerous task of manually grading student essays severely restricts teachers' ability to provide timely and effective feedback, let alone individualized guidance tailored to each learner's specific writing issues. These systemic problems lead to significant effort on the part of writing instructors, yet yield minimal educational outcomes.

Recent research has indicated that digital writing tools can exert a positive influence on learners' writing quality (Bahari, 2022; Fitria, 2023). With the advent of artificial intelligence (AI) technology in education, AI-driven writing tools, such as automated writing evaluation (AWE) systems (e.g., *Pigai* and *Criterion*), and generative AI tools have emerged as innovative solutions. AWE systems utilize natural language processing to identify errors and inappropriate expressions in learners' writing and deliver targeted feedback (Wang & Zhang, 2020). These tools not only expand practice opportunities but also alleviate teachers' grading burdens, providing personalized and immediate feedback to support learners' writing processes, enhance their skills, and foster greater autonomy (Bai & Hu, 2017).

The recent breakthroughs in highly sophisticated large language models (LLMs), such as ChatGPT and DeepSeek, are poised to revolutionize automated writing evaluation (Bai & Wei, 2024). Modern LLMs go beyond mere error correction by generating coherent and compositionally sound essays. With simple prompts, these generative AI tools can produce grammatically accurate and contextually appropriate texts. For English-as-a-Foreign-Language (EFL) writers, LLMs offer a range of benefits, including machine translation of native-language drafts into English, bulk error correction, tone adjustments, and the addition of coherence-enhancing elements such as discourse markers (Tate et al., 2023). Despite widespread agreement among scholars that educational practices must adapt to the rapid advancements in AI (Tseng & Warschauer, 2023), empirical evidence regarding the meaningful enhancement of learning outcomes through the integration of generative AI remains limited (Chiu et al., 2023). Against this backdrop, this study seeks to investigate whether AI-integrated EFL instruction can effectively improve learners' writing proficiency.

Literature Review

The burgeoning research on the application of AI technology in second language (L2) writing has predominantly converged on two major domains: writing outcomes and writing processes (Wangdi, 2022; Ng et al., 2023). Investigations into writing outcomes, which represent the most extensively explored area, have scrutinized the impact of AWE systems on learners' essay quality and overall writing proficiency. These studies typically employed within-group comparisons, examining improvements in learners' holistic scores or sub-scores (e.g., vocabulary, syntax, content) following exposure to AWE systems. Additionally, between-group comparisons were frequently utilized, with experimental and control groups designed to analyze disparities in writing skills over periods ranging from four months to a year and a half. Research on writing processes, by contrast, has delved into behavioral patterns such as the frequency of

revisions (a proxy for learner engagement), the types of feedback adopted, and the effectiveness of modifications made in response to feedback.

Despite the increasing prevalence of theoretical discussions regarding the integration of generative AI tools into EFL education (Chen, 2024; Wen & Liang, 2024), empirical studies within the realm of writing pedagogy remain relatively sparse. For instance, Dai et al. (2023) examined ChatGPT's error feedback for undergraduate writers and found it to be more comprehensible and detailed than traditional teacher feedback, although it was only partially aligned with instructor evaluations. Mizumoto and Eguchi (2023) assessed ChatGPT's performance in evaluating 12,100 non-native essays, demonstrating its reliability in rubric-based scoring. Ren et al. (2024) observed that ChatGPT provided more structural feedback and direct error corrections compared to human teachers. Bai and Wei (2024) introduced a "Composing-Comparison-Rewriting" model to examine how EFL learners integrate ChatGPT's paraphrasing feedback, revealing that such feedback enhanced attention to lexical weaknesses and revision quality, contingent upon learners' awareness of feedback quantity and type. However, these studies are notably deficient in controlled pedagogical experiments that could systematically verify whether AI integration leads to sustained improvements in writing proficiency.

Notably, existing research is marred by significant methodological limitations. Outcome studies often conflate score increases (whether total or sub-scores) with linear, "staircase-like" progression (Bai & Ye, 2018), thereby neglecting the inherently nonlinear and "web-like" nature of L2 development. Process studies, too, tend to infer behaviors from textual outputs rather than holistically observing learners' interactions with AWE systems. Since the 1990s, scholars have progressively shifted from linear models of L2 acquisition to Complex Dynamic Systems Theory (CDST; Larsen-Freeman, 1997), which reconceptualizes language as a dynamic, adaptive system characterized by (1) synchronic and diachronic evolution, (2) complexity through interconnected subsystems (phonology, lexicon, syntax), (3) nonlinear and chaotic development, (4) sensitivity to initial conditions, and (5) self-organization and feedback-driven adaptation. This theoretical framework posits that L2 competence develops not linearly but through cyclical fluctuations—progression, regression, and stagnation (Bai & Ye, 2018). In recent years, scholars have increasingly embraced CDST to reinterpret L2 learning trajectories (Dai & Zhou, 2016; Zheng, 2019; Dai & Yu, 2022).

The present study endeavors to integrate CDST with AI-enhanced college English writing instruction, thereby investigating the dynamic and individualized pathways of EFL learners' writing development. The following two research questions are to be answered:

(1) To what extent does AI-driven English writing instruction enhance the writing proficiency of EFL learners (i.e., holistic scores, writing complexity, accuracy, and fluency)?

(2) How do the developmental trajectories of EFL learners' writing proficiency evolve within an AI-driven writing environment?

Research Design

Participants and Procedure

This study was conducted during the Fall semester of the 2024–2025 academic year at a comprehensive university located in southwestern China. Utilizing a non-probability sampling method, twelve third-year students majoring in non-English disciplines (9 female, 3 male; age range: 20–22 years, M=20.37, SD=0.53) were recruited. Participants were selected based on the inclusion criterion of having participated in the College English Test Band 4 (CET-4) administered in June 2024. CET-4 is a nationwide high-stakes standardized English test, specifically designed for non-English major students. The writing scores of the CET-4 test indicated that the participants had intermediate proficiency in English writing (M=10.74/15, SD=0.66). At the time of the study, all participants had fulfilled the mandatory English courses and were no longer enrolled in any English classes. However, they had great interest in learning English and were strongly motivated to improve their English writing skills.

Research Tool: Kimi

The AI writing assistant Kimi, developed by Beijing Moonshot AI Co., Ltd. (launched October 9, 2023), was employed to provide writing feedback. Kimi, a free tool with an intuitive user interface, is recognized for its capacity to process up to 200,000 Chinese characters and is widely utilized in academic translation, legal analysis, and API documentation. Preliminary evaluations by the researchers confirmed Kimi's robust capabilities in English text generation and error correction. To ensure pedagogical alignment, Kimi was configured to deliver concise feedback based on the CET-4 scoring criteria, focusing on three key dimensions:

(1) Content Development: Assessing the clarity of the thesis statement, argumentation, and logical coherence;

(2) Organization: Evaluating structural integrity and paragraph cohesion;

(3) Language Use: Examining vocabulary accuracy, grammatical correctness, and punctuation.

Data Collection and Analysis

This study employed a longitudinal design spanning 14 weeks. In Week 1, participants received training on the functionalities of Kimi and completed a baseline writing test. From Weeks 1 to 14, participants wrote one essay per week, each with a difficulty level comparable to that of the CET-4 writing section. They revised their essays using feedback generated by Kimi during one-hour sessions and submitted their revised drafts. From Week 4 to Week 14, another six writing tests were administered biweekly to examine the dynamic development of participants' writing competence following the use of Kimi. The writing tests were derived from the authentic writing sections of CET-4 tests. The writing section of CET-4 constitutes the initial module of the examination, specifically designed to evaluate test-takers' proficiency in English written expression. Candidates are allocated 30 minutes to complete a written response within the range

of 120 to 180 words. The writing prompts are diverse in format and may encompass tasks such as describing charts or graphs, articulating viewpoints, crafting argumentative essays, writing picture-based compositions, continuing from a provided opening sentence, or developing compositions using given keywords. The topics are often centered around social issues, campus life, cultural phenomena, and other contemporary themes that are relevant to students' scholastic and societal contexts. Ethical protocols were strictly adhered to throughout the study: participants provided informed consent, received no academic incentives, and were compensated with a monetary token upon completion of the study. Two experienced CET-4 raters independently scored all tests using the official holistic rubric adopted by the CET-4 writing assessment. Inter-rater reliability was high across all tests (r>0.7).

This study also sought to investigate the influence of incorporating AI into writing instruction on the writing complexity, accuracy, and fluency (CAF) of EFL learners' essays. The procedure employed to assess the accuracy of the essays was meticulously structured as follows: Initially, two evaluators, comprising the second author of this article and an additional seasoned writing instructor, collaboratively appraised a set of 12 essays, specifically those from the initial test administration. The inter-rater reliability between these two assessors was computed to be 0.841. Any discrepancies that emerged during this initial phase were systematically resolved through thorough discussion and consensus-building. Subsequently, the essays from the remaining six test administrations were independently evaluated by the writing instructor. The accuracy score was determined through a standardized formula: An error-free essay was assigned a baseline score of 100 points, with a deduction of one point for each error identified per 100 words. The final accuracy score was subsequently derived by subtracting the total number of errors per 100 words from the baseline score of 100. In terms of lexical complexity, this current research drew on Jarvis's (2002) research and employed the U-index, an adaptation of the type/token ratio (TTR). The traditional TTR is vulnerable to fluctuations in text length, as longer compositions tend to exhibit greater lexical repetition. The U-index, however, mitigates this limitation through its unique calculation formula: (log tokens)^2 / (log tokens - log types). For syntactic complexity, this study measured T-unit length, which was determined by dividing the total number of words by the number of T-units. Words per minute (WPM) served as the operational measure for evaluating writing fluency (Tabari et al., 2024).

Upon completion of data collection, the researchers employed SPSS (version 27) to perform inter-rater reliability tests and enable systematic comparison of mean scores and CAF indices between Test 1 and Test 7. Besides, to visualize developmental trends, dynamic line graphs were generated using Microsoft Excel, offering dual analytical perspectives: (1) Group-level trajectories, highlighting aggregate patterns in writing proficiency across the intervention period; and (2) Individual-level variability: Capturing divergent learner pathways, which contextualize group findings.

Results and Discussion

Effects of AI-driven Instruction on EFL Learners' Writing Proficiency

To determine whether the leverage of generative AI in English writing instruction has a significant impact on learners' writing proficiency, we compared learners' initial (Test 1) and final (Test 7) states of English

writing achievements. The Shapiro-Wilk test results indicated that the *W* values for all groups ranged from 0.946 to 0.987, with all p-values exceeding 0.05, suggesting that the data sets were normally distributed. Thus, paired-samples *t* test was performed to discern the within-group differences in the wiring outcomes of Test 1 and Test 7. As shown in Table 1, learners' overall writing proficiency, writing complexity and accuracy significantly improved over time ($p<0.01$) with large effect sizes (Cohen's $d>0.8$). However, there was no significant difference between Test 1 and Test 7 in writing fluency ($p>0.05$), despite a slight improvement in this dimension. Given that these students no longer had access to formal English courses, their sole engagement with English writing occurred during their weekly participation in this study. Thus, these findings suggest that employing Kimi as an English writing tutor can effectively enhance writing proficiency to a notable degree.

Table 1. Results of paired-samples *t* test

Variables	MD (T1-T7)	SD	SE	95% Confidence Interval		t	p	Cohen's d
				Lower Bound	Upper Bound			
Writing Score	-2.25	0.622	0.179	-2.645	-1.855	-12.539	.000	1.94
Syntactic Complexity	-5.126	2.58	0.745	-6.766	-3.487	-6.881	.000	2.85
Lexical Complexity	-4.797	3.928	1.134	-7.293	-2.302	-4.231	.001	1.77
Accuracy	-8.27	4.489	1.296	-11.122	-5.418	-6.382	.000	1.73
Fluency	-0.31	0.884	0.255	-0.871	0.252	-1.214	.250	0.32

MD, mean difference; SE, standard error.

The results of our research align with those of previous studies (e.g., Tsai et al., 2024; Zhao, 2024), suggesting that the integration of GenAI into writing instruction is conducive to promoting learners' writing proficiency. The research findings also indicate that AI-enabled writing instruction potentially exerts a pronounced positive influence on lexical complexity, syntactic complexity, and accuracy in English writing (Feng, 2025). In the realm of lexical complexity, AI tools can furnish a wealth of lexical resources and collocation suggestions, thereby empowering learners to deploy more precise and diverse vocabulary. For example, AI writing assistants are capable of recommending advanced vocabulary based on the contextual nuances, which significantly enhances the lexical richness and professionalism of compositions. Concerning syntactic complexity, AI tools can generate sophisticated sentence structures and subordinate clauses, facilitating learners in constructing more elaborate and logically coherent texts. Empirical research has demonstrated that AI-generated texts exhibit a higher degree of syntactic complexity compared to those produced by average learners, thereby effectively improving the organizational structure and depth of writing (Kim et al., 2025). In terms of accuracy, AI tools enhance the standardization of writing through advanced grammar-checking and spell-correction functions. Tools such as Grammarly can promptly identify and rectify grammatical errors, ensuring textual accuracy and consistency. These features not only elevate the overall quality of writing but also bolster learners' confidence in their writing capabilities. Learners' improvement in writing fluency was not statistically significant. This could be attributed to the nature of the CET writing section, which imposes both a time

constraint and a specific word limit. During the test, learners likely adhered strictly to these requirements, and upon meeting the minimum threshold, they may have refrained from further production.

Developmental Trajectories of Learners' Writing Proficiency in the AI-driven Learning Context

Figure 1 vividly illustrates the dynamic changes of learners' writing performance across four key dimensions: average writing scores, complexity (lexical and syntactic), accuracy, and fluency. Given the methodological diversity in computing these indices and the inherent heterogeneity of the data, standardization was a critical preliminary step. To address this, we applied SPSS 27.0 to transform raw scores into standardized Z-scores, adopting the methodological framework proposed by Larsen-Freeman (2006).

The analysis reveals that over the 14-week period using Kimi, the 12 participants' average writing scores did not follow a linear trajectory. Instead, their performance exhibited fluctuations: a sharp rise in the first four tests, followed by a decline in Test 5, and a subsequent moderate recovery. Breaking down the trends by dimension reveals that writing accuracy demonstrated a near-linear improvement, with a gradual initial rise followed by a steeper increase in later stages. Lexical complexity remained relatively stable across the first six tests but surged significantly in the final assessment. Syntactic complexity showed a modest gain between Tests 1 and 2, then declined steadily until Test 5, before rebounding sharply in the last two tests. Fluency displayed the most volatile pattern: it peaked early (Test 2), plummeted to its lowest point by Test 4, stabilized in Tests 4-5, rebounded in Test 6, and ultimately settled slightly above baseline in Test 7.

These findings confirm the tenets of the CDST. According to de Bot (2008), language development is a nonlinear emergent process, encompassing both bottom-up construction (from local to global) and top-down constraints (from global to local). As a complex system, it demonstrates full interconnectedness where all constituent dimensions are interrelated; a change in one component may trigger cascading effects on others and ultimately reshape the entire system's behavior, epitomizing the holistic principle that moving one part affects the whole. In essence, the CDST rejects simplistic linear causality. A system's behavior emerges from the collective configuration of all its components—meaning all relevant factors operate synergistically (Dörnyei, 2014). Crucially, the development of the components is often nonlinear. As can be seen from Figure 1, learners' writing CAF experienced substantial fluctuations.

Besides, Figure 1 shows that the developmental trajectories of CAF dimensions exhibited nonlinear interdependencies across the seven tests, revealing dynamic trade-off relationships among these competences. For example, in Test 5, accuracy was at the highest level, followed by lexical complexity, fluency, and syntactic complexity. In contrast, in Test 7, syntactic complexity climbed to the second place, while fluency dropped to the lowest level. As posited by Vercellotti (2017), the trade-off effect constitutes an inevitable phenomenon in L2 development. This dynamic arises primarily due to limited attentional resources, which constrain learners' capacity to concurrently process all dimensions of CAF. Consequently, when learners prioritize one CAF component (e.g., syntactic complexity), their performance in another (e.g., lexical accuracy) may be compromised, reflecting the competitive relationship among these constructs (Skehan, 1998).
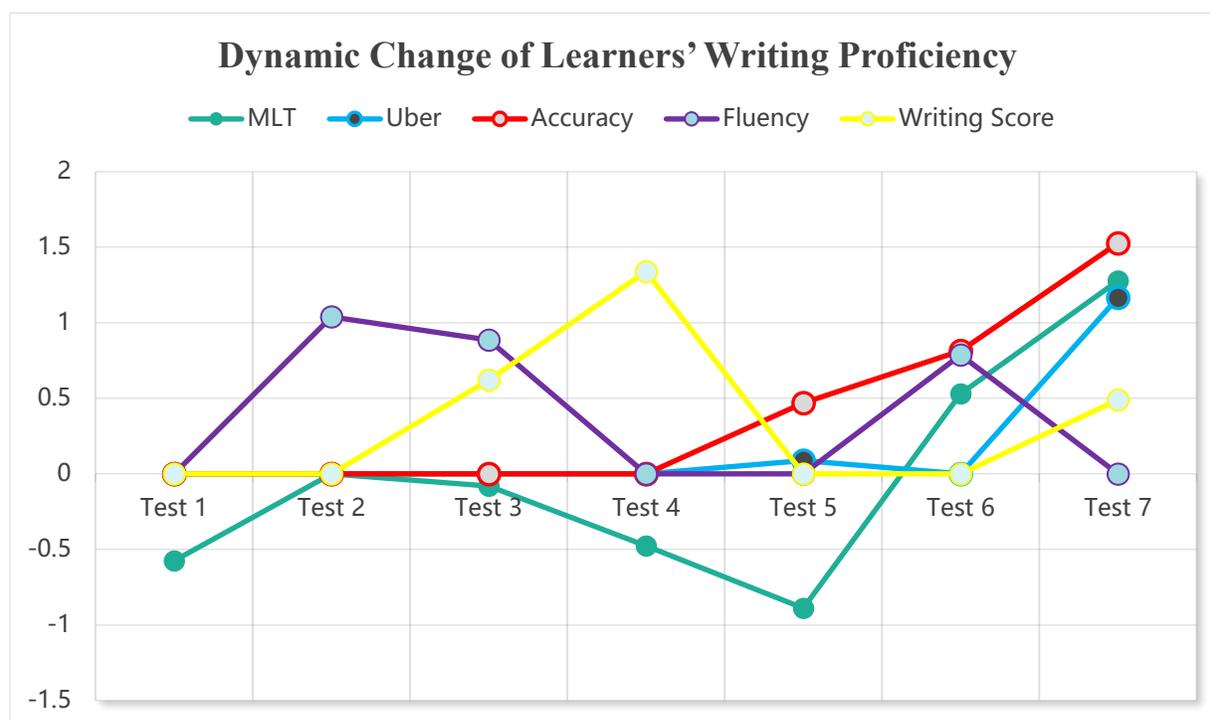
Figure 1. Dynamic changes of the average writing scores and the CAF dimensions in each test

We also compared the individual variability with regard to learners' developmental trajectories in English writing proficiency. Given the fact that it is not possible to put the pathways of all the 12 participants in each line graph, we only reported the trajectories of four representative learners (with diverse developmental patterns). Figure 2 delineates the individual score trajectories of each student from the first to the seventh test. As depicted, the overarching trend for each student's performance is upward, echoing the general trajectory of the average scores illustrated in Figure 1. However, the figure also underscores the considerable variability in writing performance among individual students. Their scores do not invariably follow a linear upward path; instead, they often exhibit fluctuations, and in some cases, even a general decline. For example, Student 8's scores display a predominantly downward trend throughout the testing period. This observation suggests that the utilization of artificial intelligence tools may not uniformly exert a positive influence on the writing proficiency of all students.
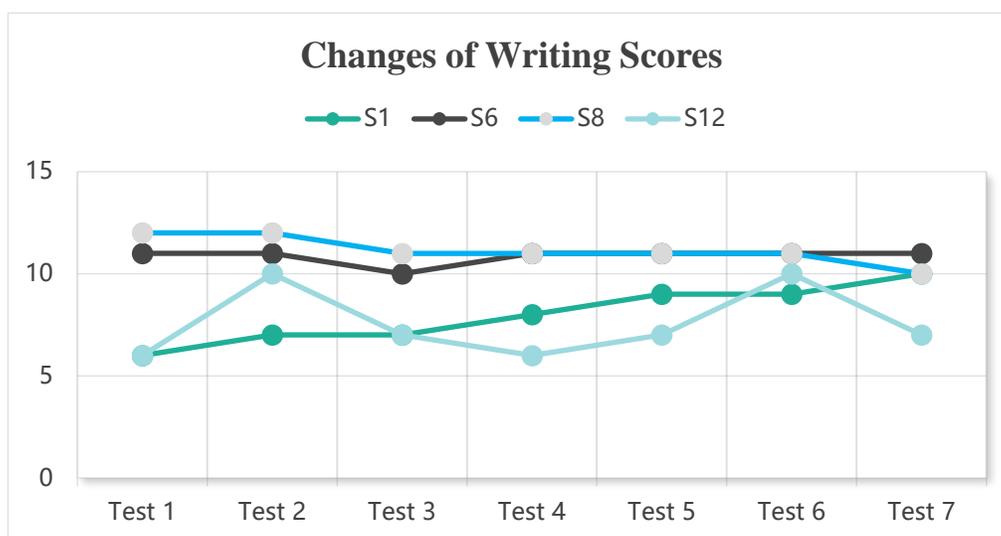
Figure 2. Dynamic changes of students' essay scores in each test

Figure 3 shows the dynamic evolution of lexical complexity in students' compositions. Overall, a general upward trajectory in lexical complexity is observed among the students. However, the developmental paths of individual students in this dimension diverge significantly, with the curves of lexical complexity displaying notable undulations and variability.
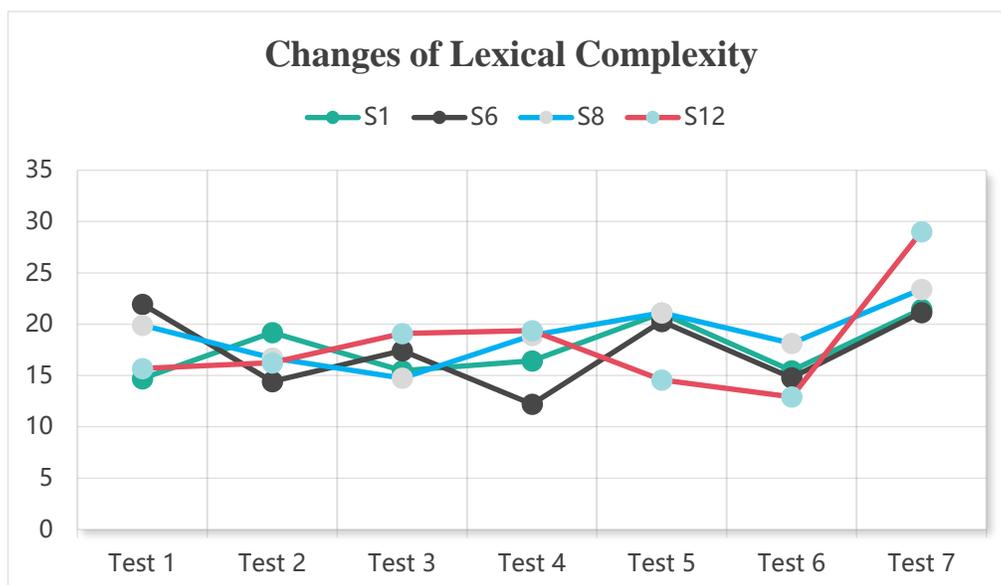

Figure 3. Dynamic changes in lexical complexity of each student's essays

As evidenced in Figure 4, while significant improvements in syntactic complexity were observed between learners' initial and final states, the developmental trajectories in this dimension exhibited marked nonlinearity, characterized by pronounced fluctuations throughout the intervention period.
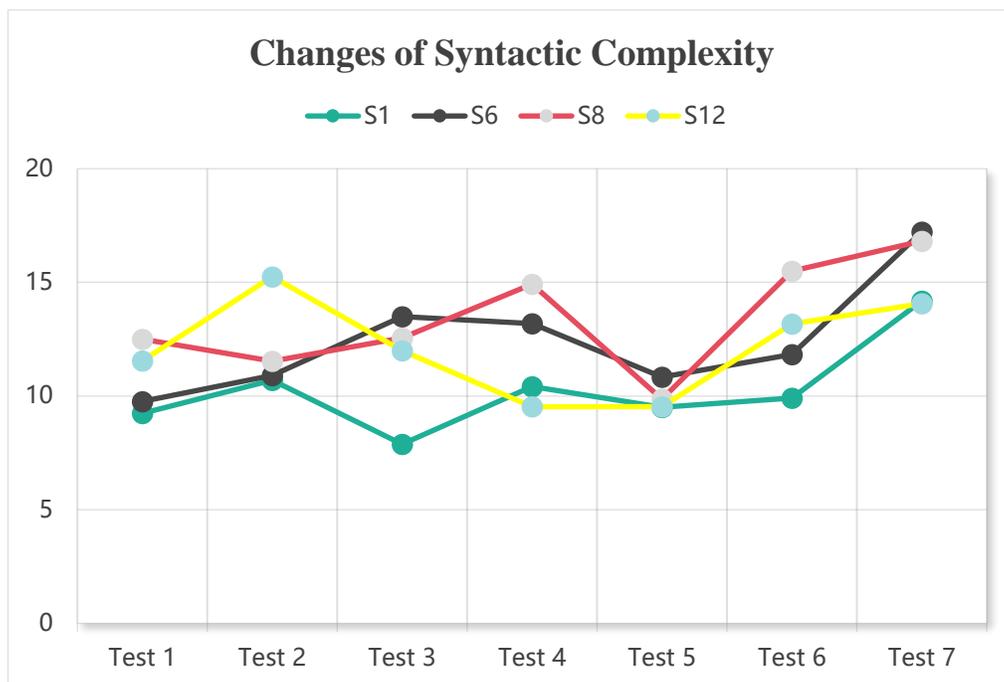

Figure 4. Dynamic changes in syntactic complexity of each student's essays

Figure 5 illustrates the dynamic change of accuracy within each student's compositions. The results indicate a consistent upward trajectory in writing accuracy among all participants, with minimal fluctuations, mirroring the overarching trend observed.
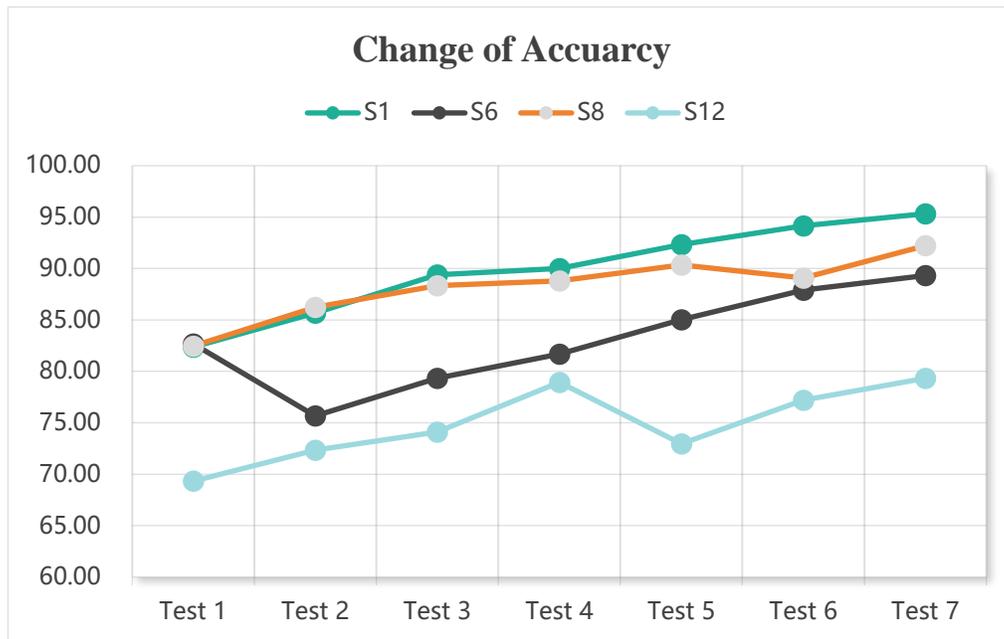
Figure 5. Dynamic changes in the accuracy of each student's essays

Figure 6 demonstrates that learners' developmental trajectories exhibited considerable fluctuations throughout the study period. The analysis reveals that while Students 1, 8, and 12 ultimately achieved writing fluency levels marginally higher than their baseline, Student 6's final performance remained slightly below the initial state.
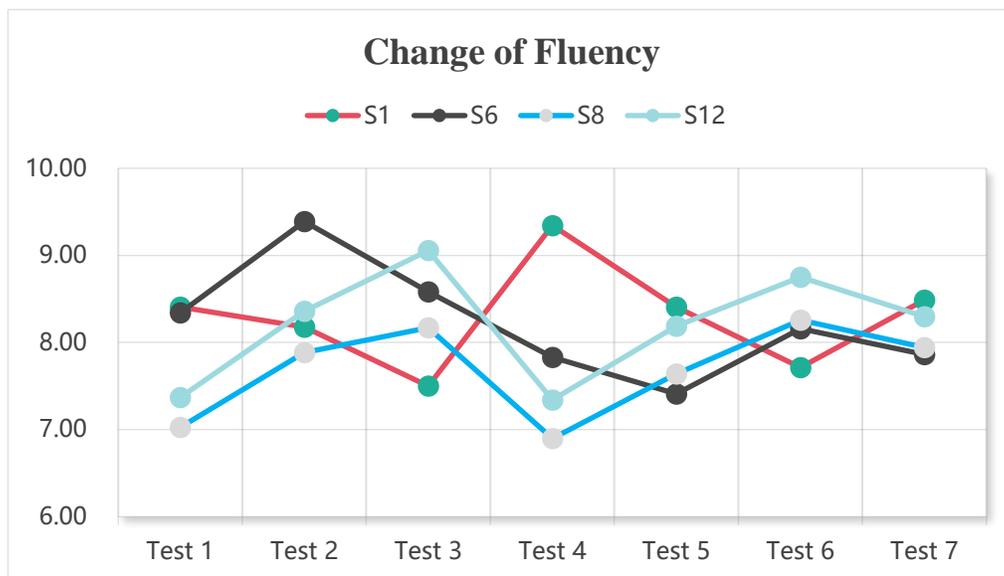

Figure 6. Dynamic changes in fluency of students' essay scores

The findings of this study further corroborate the CDST. Specifically, with regard to developmental trajectories, the writing performance of students—whether assessed in terms of accuracy or lexical and syntactic complexity—exhibited diverse patterns. Some students' performance remained relatively stable throughout the study period, while others experienced a single significant fluctuation. A subset of students demonstrated multiple fluctuations in their writing proficiency. Regarding the developmental trends, the majority of students exhibited an upward trajectory, although instances of stability or decline were also observed. In terms of the progression patterns, some students showed steady improvement over time, while others experienced a sharp rise to peak performance, followed by abrupt declines and repeated oscillations. In contrast, some students followed a spiraling upward trajectory, characterized by periods of advancement interspersed with temporary setbacks. These conclusions are consistent with the research of Larsen-Freeman (2006). Language learning outcomes emerge from the dynamic interplay of multiple factors. While individual differences in these factors may appear insignificant in isolation, their combined effects are not merely additive but rather generate complex, non-linear forces that lead to marked divergence in learners' behaviors and outcomes.

Pedagogical Implications

The findings of this study yield several salient implications for English writing instruction, particularly in the context of tertiary-level EFL education.

Firstly, the necessity of differentiated instruction is underscored by the study's results. While overall writing proficiency exhibited an upward trajectory, significant disparities in the development of lexical and syntactic complexity were observed among students. This highlights the imperative for writing teachers to recognize and accommodate individual differences by implementing personalized teaching strategies with the assistance of AI tools. Through a meticulous evaluation of each student's strengths and weaknesses, teachers can establish tailored learning objectives and devise targeted practice activities, thereby more effectively catalyzing writing development.

Secondly, the critical role of dynamic assessment and continuous feedback in enhancing students' English writing skills is evident. The longitudinal research design elucidates the dynamic and evolving nature of writing proficiency, suggesting that ongoing assessments are essential for monitoring student progress. Regular writing tasks, coupled with timely and constructive feedback, not only facilitate the prompt correction of errors but also bolster students' confidence and intrinsic motivation. Moreover, teachers should actively encourage self-assessment and peer review practices to foster students' autonomous learning capabilities and critical thinking skills.

Thirdly, writing instructors should prioritize vocabulary acquisition and syntactic instruction. The observed upward trends in lexical and syntactic complexity confirm their central importance in writing proficiency. Teachers should integrate vocabulary and syntax as core components of the writing curriculum, employing a diverse array of pedagogical approaches—such as vocabulary games and sentence restructuring exercises—to help students expand their lexical resources and master complex

syntactic structures. The strategic use of corpus tools and AI-powered writing assistants can further enrich language input and provide instant feedback, thereby accelerating linguistic development.

Lastly, the integration of technological tools with student autonomy is essential. The study demonstrates AI's positive impact on enhancing writing complexity and accuracy, suggesting that teachers should incorporate AI writing aids into classroom practices to support independent learning and writing practice. Simultaneously, teachers can guide students in the judicious use of these tools, avoiding over-reliance while nurturing their capacity for independent thinking and self-directed learning. This balanced approach ensures that technology serves to enhance, rather than supplant, students' intrinsic writing development.

Conclusion

This study employed a longitudinal research design to examine the dynamic development of English writing proficiency among 12 tertiary-level EFL learners. The findings indicate that overall writing proficiency, as well as the CAF dimensions, exhibited a general upward trend. However, the extent of improvement in these dimensions varied significantly among individual students. This study furnishes robust empirical evidence for EFL writing pedagogy, underscoring the pivotal roles of differentiated instruction, dynamic assessment, and technology-assisted tools in augmenting students' writing proficiency. Writing instructors are encouraged to adapt their pedagogical strategies with flexibility, drawing upon these insights to foster the holistic development of each learner.

However, it is important to acknowledge that the findings of this study are limited by the constraints of a small sample size and the absence of a control group. Future research should consider these limitations to provide a more comprehensive understanding of how learners' writing outcomes are influenced in the context of AI-driven writing instruction.

Funding Statement

References

Bahari, A. (2023). Affordances and challenges of technology-assisted language learning for motivation: A systematic review. *Interactive Learning Environments*, 31(9), 5853-5873.

Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond?. *Educational Psychology*, 2017, 37(1), 67-81.

Bai, L., & Wei, Y. (2024). Exploring EFL learners' integration and perceptions of ChatGPT's text revisions: A three-stage writing task study. *IEEE Transactions on Learning Technologies*, 17, 2161-2172.

Bai, L., & Ye, S. (2018). The dynamic development of English-L2 writing competence. *Modern Foreign Languages*, 41(3), 354-366.

Chen, X. (2024). ChatGPT and second language pragmatic teaching. *Foreign Language Learning Theory and Practice*, 4, 16-29.

Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, 100118.

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gašević, D., & Chen, G. (2023, July). 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.

Dai, Y., & Yu, H. (2022). A retrodictive study of second language motivational dynamics: A Complex Dynamic Systems perspective. *Modern Foreign Languages*, 45(3), 357-368.

Dai, Y., & Zhou, L. (2016). Study of second language acquisition from the perspective of the Dynamic Systems Theory: Limitations and countermeasures. *Foreign Language World*, 74(3), 57-63, 96.

de Bot, K. (2008). Second language development as a dynamic process. *The Modern Language Journal*, 92(2), 166-178.

Dörnyei, Z. (2014). Researching complex dynamic systems: 'Retrodictive qualitative modeling' in the language classroom. *Language Teaching*, 47(1), 80-91.

Feng, L. (2025). Investigating the effects of artificial intelligence-assisted language learning strategies on cognitive load and learning outcomes: A comparative study. *Journal of Educational Computing Research*, 62(8), 1961-1994.

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essays. *ELT Forum: Journal of English Language Teaching*, 12(1), 44-58.

Jarvis, S. (2002). Short texts, best-fitting curves, and new measures of lexical diversity. *Language Testing*, 19(1): 57-84.

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies*, 30(1), 1265-1300.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141-165.

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, 2(2), 100050.

Ng, D. T. K., Leung, J. K. L., Su, J., Ng, R. C. W., & Chu, S. K. W. (2023). Teachers' AI digital competencies and twenty-first century skills in the post-pandemic world. *Educational Technology Research and Development*, 71(1), 137-161.

Ren, W., Liu, Y., & Xie, Y. (2024). A comparative study of ChatGPT and teacher feedback in English writing instruction. *Foreign Language Learning Theory and Practice*, 4, 30-38, 60.

Skehan, P. (1998). A cognitive approach to language learning. Oxford University Press.

Tabari, M. A., Khajavy, G. H., & Goetze, J. (2024). Mapping the interactions between task sequencing, anxiety, and enjoyment in L2 writing development. *Journal of Second Language Writing*, 65, 101116.

Tate, T., Doroudi, S., Ritchie, D., Xu, Y., & Warschauer, M. (2023). Educational research and AI-generated writing: Confronting the coming tsunami. *EdArXiv*, 10.

Tsai, C. Y., Lin, Y. T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, 1-19.

Tseng, W., & Warschauer, M. (2023). AI-writing tools in education: If you can't beat them, join them. *Journal of China Computer-Assisted Language Learning*, 3(2), 258-262.

Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics*, 38(1), 90-111.

Wang, J., & Zhang, T. (2020). A study of the relationship between the textual quantitative indices of L2 writing and machine scores. *Foreign Language Testing and Teaching*, 3, 12-20.

Wangdi, T. (2022). Student teachers ' perception of academic writing: challenges and pedagogical needs. *Journal of Education for Teaching*, 48(3), 378-380.

Wen, Q., & Liang, M. (2024). Human-AI interactive negotiation competence: ChatGPT and foreign language education. *Foreign Language Teaching and Research*, 56(2), 286-296, 321.

Zhao, D. (2024). The impact of AI-enhanced natural language processing tools on writing proficiency: An analysis of language precision, content summarization, and creative writing facilitation. *Education and Information Technologies*, 1-32.

Zheng, Y. (2019). The CDST Perspective on Effective Foreign Language Teaching. *Contemporary Foreign Language Studies*, 443(5), 12-16, 49.