

RESEARCH IN PROGRESS

AMERICAN ASSOCIATION FOR APPLIED LINGUISTICS CONFERENCE, HOUSTON 2024
SELECTED POSTER PRESENTATIONS*

Exploring ChatGPT's potential as an AI-powered writing assistant: A comparative analysis of second language learner essays

Shuyuan Tu 

Georgia State University, Atlanta, USA
Email: stu2@gsu.edu

(Received 1 June 2024; accepted 13 June 2024)

1. Introduction

The emergence of advanced artificial intelligence (AI) technology has led to the development of innovative writing tools, which are valuable for second language (L2) writers and language learners (Alharbi, 2023). Large language models (LLMs), like GPT-3.5 and GPT-4.0, demonstrate robust capabilities in generating human-like texts, making them accessible tools for writers (Alharbi, 2023; Eaton et al., 2021). While previous research has investigated applications of ChatGPT in supporting and evaluating writing, such as engaging learners in discussions about writing prompts and the writing process (Hartwell & Aull, 2022), providing feedback on language use (Su et al., 2023), and synthesizing literature (Dowling & Lucey, 2023), its potential as an editing tool for enhancing linguistic complexity in L2 writing remains unexplored.

Linguistic complexity, including lexical and syntactic complexities, is a critical construct in L2 writing development and assessment (Lu, 2012, 2017; Ortega, 2003). Previous studies have explored different indices of linguistic complexity as measures in L2 writing development (Ortega, 2003). Understanding the relationship between linguistic complexity and learners' proficiency and writing quality in L2 writing is important for investigating how AI-powered writing assistants can support L2 writing.

1.1 Research questions

This study aims to investigate the potential of ChatGPT as an AI-powered writing assistant in offering editing feedback to enhance linguistic complexity in L2 learner essays. Specifically, it examined the performance of ChatGPT in improving lexical and syntactic complexity and evaluated its reliability compared with human editors by addressing the following research questions:

1. How are the lexical complexity measures of ChatGPT-edited learner essays different from those of learner essays?
2. How are the syntactic complexity measures of ChatGPT-edited learner essays different from those of learner essays?
3. To what extent does ChatGPT as an AI-powered writing assistant contribute to linguistic complexity in comparison with human editors?

*A reproduction of the poster discussed is available in the supplementary material published alongside this article on Cambridge Core.

2. Method

2.1 Data collection

The data for this study consisted of 140 essays written by high-intermediate level learners extracted from the International Corpus Network of Asian Learners of English (ICNALE; Ishikawa, 2018), along with their corresponding human-edited and ChatGPT-edited versions.

2.1.1 Learner essays and human-edited essays

The learner essays were extracted from the ICNALE Written Essays submodule, and their human-edited versions were obtained from the ICNALE Edited Essays submodule. The human-edited essays were fully edited versions of learner essays, edited by five professional editors who used the ESL Composition Profile rubric (Jacobs, 1981).

2.1.2 ChatGPT-edited essays

ChatGPT was given the following prompt: “Can you edit L2 learner essays at B2+ level based on the ESL Composition Profile (Jacobs, 1981), which uses five rating criteria: Content (CON), Organization (ORG), Vocabulary (VOC), Language use (LNU), and Mechanics (MEC).” ChatGPT was then required to generate fully edited versions of the 140 learner essays using the GPT-3.5 API in Python. The generated essays were manually double-checked by human to ensure that ChatGPT followed the rubric.

2.2 Data analysis

The study adopted a mixed-method approach in its data analysis. For the quantitative analysis, 25 indices to evaluate lexical density, sophistication, and variation were measured using the *Lexical Complexity Analyzer* (Ai & Lu, 2010; Lu, 2012), while 15 indices of syntactic complexity, including the length of production units, amount of subordination and coordination, and degree of phrasal sophistication, were measured using the *L2 Syntactic Complexity Analyzer* (Ai & Lu, 2013; Lu, 2017). One-way ANOVAs and MANOVAs were conducted to investigate the differences in lexical and syntactic complexity measures between learner essays, ChatGPT-edited essays, and human-edited essays.

For the qualitative analysis, a sample of 20 essays from each corpus (a total of 60 essays) was randomly selected. The qualitative analysis involved a close examination and comparison of the specific editing moves and changes made by human editors versus ChatGPT in relation to lexical and syntactic complexity within this.

3. Results and discussion

The quantitative analysis revealed that ChatGPT-edited essays showed significant differences in lexical density and variation ($p < 0.001$) compared with the learner essays, but no significant changes in lexical and verb sophistication. At the syntactic level, ChatGPT-edited essays showed significant differences in the use of coordinating conjunctions ($p < 0.001$) compared with learner essays, but no significant differences in length of production units, subordination, or phrasal sophistication.

The qualitative analysis highlighted differences in editing moves between ChatGPT and human editors. Compared to human editors, ChatGPT tended to replace words with more sophisticated alternatives or diverse lexical choices, while human editors focused on correcting misspelled words without altering them significantly. Additionally, ChatGPT tended to use more coordinating conjunctions (e.g., “and,” “or”), whereas human editors preferred to maintain the original sentence format. Regarding phrasal sophistication, ChatGPT tended to replace single verbs with phrases (e.g., replacing “mind” with “take into account,” “died” with “led to death”), while human editors retained most of the original phrasal structure.

Overall, the findings suggest that ChatGPT could contribute to increased lexical variation, a broader lexical repertoire, and higher numbers of coordinating conjunctions in editing learner essays.

Moreover, the study reveals differences between the editing performed by ChatGPT and human editors, with ChatGPT providing a range of alternatives potentially enhancing lexical diversity and different dimensions of syntactic complexity, while human editors lean towards conserving the learner's original expression and complexity.

4. Conclusion and limitations

This study explored the potential of using ChatGPT as an AI-powered writing assistant with a focus on its editing performance in contributing to linguistic complexity in L2 learner essays. The findings underscore the potential role that generative AI can play in language learning and the importance of balancing technological assistance with the preservation of authorial integrity in L2 writing.

While this study provides preliminary insights, further research is needed to investigate the reliability of ChatGPT in making consistent editing decisions across various prompts and a larger corpus of learner texts. Future studies may continue to explore the integration of generative AI into language teaching and learning, particularly as a support mechanism for L2 writing.

Supplementary material. The supplementary material for this article can be found at: <https://doi.org/10.1017/S0261444824000259>.

References

- Ai, H., & Lu, X. (2010). A web-based system for automatic measurement of lexical complexity. In *27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10)*. Amherst, MA. June (pp. 8–12).
- Ai, H., & Lu, X. (2013). A corpus-based comparison of syntactic complexity in NNS and NS university students' writing. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (Eds.), *Automatic treatment and analysis of learner corpus data* (pp. 249–264). Amsterdam, the Netherlands.
- Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated writing assistance tools. *Education Research International*, 2023(1), 1–15. <https://doi.org/10.1155/2023/4253331>.
- Dowling, M., & Lucey, B. (2023). ChatGPT for (finance) research: The Bananarama conjecture. *Finance Research Letters*, 53, 103662. <http://doi.org/10.1016/j.frl.2023.103662>.
- Eaton, S. E., Mindzak, M., & Morrison, R. (2021). *Artificial intelligence, algorithmic writing & educational ethics [Paper Presentation]*. Canadian Society for the Study of Education Société canadienne pour l'étude de l'éducation, Edmonton, AB, Canada.
- Hartwell, K., & Aull, L. (2022). Constructs of argumentative writing in assessment tools. *Assessing Writing*, 54. <https://doi.org/10.1016/j.asw.2022.100675>.
- Ishikawa, S. (2018). The ICNALE edited essays: A dataset for analysis of L2 English learner essays based on a new integrative viewpoint. *English Corpus Studies*, 25, 117–130.
- Jacobs, H. L. (1981). *Testing ESL composition: A practical approach. English composition program*. Newbury House Publishers, Inc.
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *Modern Language Journal*, 96(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x.
- Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, 34(4), 493–511. <https://doi.org/10.1177/0265532217710675>.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492–518. <https://doi.org/10.1093/applin/24.4.492>.
- Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, 57, 100752. <https://doi.org/10.1016/j.asw.2023.100752>.