

Using AI Chatbots to Improve University EFL Students' Speech Scripts' Writing Quality and Critical Thinking Skills

Meng Zhang^{1,*}, Yuxuan Wang¹ and Qian Shang¹

¹School of English Studies, Sichuan International Studies University, China

Abstract

The purpose of this study was to investigate effects of artificial intelligence (AI) chatbots on university English as a foreign language (EFL) students' language use and critical thinking skills in English speech script writing. The 18-week English-speaking course was designed and conducted on the basis of production-oriented approach, a theory of foreign language education with Chinese features. By integrating the production-oriented approach with the technological capabilities of the chatbot, the course instructor offered 30 university EFL students theoretical instruction and task-driven writing activities focused on public speaking in English. Throughout the course, the participants worked in collaboration with the AI chatbot to complete each script writing assignment. Ultimately, the instructor collected and analyzed all original drafts, revised drafts, and transcripts of conversations between the students and the AI chatbot. The study's results showed that: (1) four indicators of lexical complexity in the revised drafts were significantly improved compared with the original drafts ($p < 0.01$), whereas one indicator of syntactic complexity also demonstrated significant enhancement ($p < 0.05$); (2) in comparison with the original drafts, the majority of the revised drafts (71.93%) showed modifications in regard to claims, warrants, and grounds; (3) the contents of the conversations between students and AI chatbot were grouped into three themes: making suggestions for language revisions, extending and developing the main points, and providing supporting

*Corresponding Author E-mail: zhang.meng@sisu.edu.cn

materials for arguments. The study's results can be taken as a reference by English teachers for designing and conducting writing and speaking courses. The study also contributes empirical evidence towards building AI-assisted EFL teaching theories and methods in the era of AI.

Keywords: AI chatbot, L2 writing, Critical thinking, English public speaking

The integration of artificial intelligence (AI) chatbot-assisted learning in English as a foreign language (EFL) education has become a prominent trend within the realm of computer-assisted language learning. Recent scholarly work has underscored the significant potential of AI chatbots in enhancing the English writing process and improving the quality of written output. Empirical studies have demonstrated that AI chatbots can effectively fulfill the roles of tutor, collaborator, and language output facilitator for L2 writers (Barrot, 2023; Guo et al., 2022; Su et al., 2023). The capabilities of AI chatbots, such as error correction, facilitation the generation of ideas, and the provision of timely and adaptive feedback, have been widely acknowledged and integrated into L2 writing instruction.

However, the literature has also revealed certain limitations of AI chatbots, particularly in the areas of logical reasoning (Zhang & Li, 2021; Thorp, 2023), deep cohesion (Zhou et al., 2023), and the possibilities for generating inauthentic or misleading responses (OpenAI, 2022). These shortcomings have raised significant concerns that over-reliance on AI tools could diminish students' critical thinking and creativity. Barrot (2023) has highlighted the risk of students becoming overly dependent on AI for the sake of convenience.

Despite these concerns, there is a notable gap in the research regarding the impact of AI chatbots on the critical thinking skills of EFL students. The extent to which AI chatbots influence the critical thinking of L2 learners remains an open question. In response to the pressing need to prioritize the development of critical thinking within L2 writing pedagogy, this study addressed a dual objective: to enhance students' English language proficiency and to augment their critical thinking abilities. To achieve this, we incorporated an AI chatbot into an English-speaking course for 30 Chinese university students and conducted an assessment that evaluated both the quality of students' speech script writing and the application of critical thinking skills in their writing process.

This study aimed to contribute to the existing body of knowledge by providing empirical insights into the interplay between AI chatbot-assisted learning and the cultivation of critical thinking in EFL contexts.

Overview of the Methodology

The current study used an action research methodology as outlined by Kemmis et al. (2014). The process adheres to a cyclical pattern of actions: planning, acting, observing, reflecting, and revising.

The 18-week course aims to equip students with the ability to independently draft and eloquently present public speaking scripts in English. It uses the product-oriented approach (Wen, 2017; 2018), which aligns pedagogical tasks with the course objectives and emphasizes language output. The instructional model is structured around three key phases: initiation, facilitation, and evaluation, with a particular focus on the instructor's role in guiding and assessing student progress. The curriculum also integrates AI chatbots to enhance interactivity, making them provide tailored feedback on task-related content for learners. This design effectively merges the pedagogical strategies of product-oriented approach with the dynamic engagement offered by AI chatbots.

The course instructor, who is also the principal investigator of this research project, offered 30 university EFL students foundational theoretical instruction and task-driven writing activities focused on public speaking in English. Throughout the course, participants were working in collaboration with the AI chatbot to complete each script writing assignment. Ultimately, the instructor collected and analyzed all original drafts, revised drafts, and transcripts of conversations between the students and the AI chatbot.

Representative Results

Changes in Lexical and Syntactic Complexity

To uncover the shifts in writing quality, we examined the lexical and syntactic complexity of all collected drafts and compared the differences between the original and revised versions. The indices and measurement methods for assessing lexical and syntactic complexity are detailed in Table 1, with the findings presented in Table 2.

In terms of lexical complexity, our analysis revealed that the use of the 1000 most common words in the revised drafts significantly decreased compared with their use in the original drafts ($t = 2.877, p < 0.01$). Conversely, the use of academic vocabulary ($t = -3.084, p < 0.01$), the type-token ratio ($t = -8.767, p < 0.01$), and lexical density ($t = -3.734, p < 0.01$) all significantly increased. Collectively, these four indices suggest that the lexical complexity of the revised drafts was substantially enhanced following the AI chatbot's intervention.

Table 1. The indices and measurements of lexical and syntactic complexity

	Indices	Concepts	Measurements
Lexical complexity	Most commonly used 1000 words	The 1000 most frequent words in the Lexical Frequency Profile corpus	Lexical frequency profile (http://www.lex tutor.ca/vp/) (Cobb, 2002; Heatley et al., 2002)
	Academic words	Words in Coxhead's Academic Word List (Coxhead, 2000)	
	TTR	Type–token ratio	Number of different words/total number of words
	Lexical density	The ratio of content words to total vocabulary in a text	Content words/total
Syntactic complexity	Word count	Total number of words in a text	NeoSCA 0.1.2 (computer software) (Lu, 2010; 2012; Tan, 2024)
	MLS	Mean length of sentence	
	MLT	Mean length of T-unit	
	MLC	Mean length of clauses	

Table 2. Results of paired-samples t-tests of changes in lexical and syntactic complexity (N = 57)

	Original Draft		Revised Draft		MD	t	Sig. (Two-Tailed)
	M	SD	M	SD			
Most commonly used 1000 words (n = 57)	77.99%	4.11%	76.82%	4.82%	1.17%	2.877	<0.01
Academic words (n = 57)	5.89%	2.52%	6.56%	2.73%	−0.67%	−3.084	<0.01
Type-token ratio (n = 57)	0.46	0.62	0.52	0.43	−0.063	−8.767	<0.01
Lexical density (n = 57)	0.56	0.43	0.57	0.40	−0.01	−3.734	<0.01
Word count (+) (n = 24)	281.88	88.19	399.58	71.41	−117.70	−5.656	<0.01
Word count (−) (n = 33)	444.27	136.47	347.45	61.99	96.82	5.108	<0.01
MLS (n = 57)	17.04	2.65	16.38	2.15	0.66	2.048	<0.05
MLT (n = 57)	16.35	2.15	15.89	2.29	0.46	1.551	0.127
MLC (n = 57)	11.71	1.57	11.79	1.65	−0.08	−0.434	0.666

As for syntactic complexity, we assessed and reported the three indices, namely the mean length of sentences (MLS), mean length of T-units (MLT), and the mean length of clauses (MLC) in Table 2. We discovered that the MLS ($t = 2.048$, $p < 0.05$) of the revised drafts was statistically higher than that of the original drafts, indicating a significant increase in sentence length. However, the MLT and MLC of the revised drafts did not exhibit significant changes when compared with the original drafts.

Changes in Critical Thinking Skills

The Toulmin Model (Toulmin, 1958) was used to evaluate the critical thinking skills exhibited by students in script writing. We constructed an analytical framework based on the Toulmin Model, which included three primary elements: claims, warrants, and grounds. Each primary element was further divided into two or three sub-elements.

Two researchers meticulously read each draft, marked a total of 130 changes between the original and revised versions, and then categorized these changes into eight distinct categories within the analytical framework, aligning with the concept of each element.

Figure 1 illustrates that, out of the 57 drafts, 41 (71.93%) of them demonstrated changes that could be fitted into the analytical framework, whereas 16 (28.07%) did not exhibit the necessary changes. Among the total modifications ($n = 130$), we identified that 49 (37.69%) modifications related to claims, 29 (22.31%) related to warrants, and 52 (40%) related to grounds.

Figure 2 provides a more detailed breakdown: in the claim category, 13 out of 130 modifications (10%) either added or reduced the number of claims,

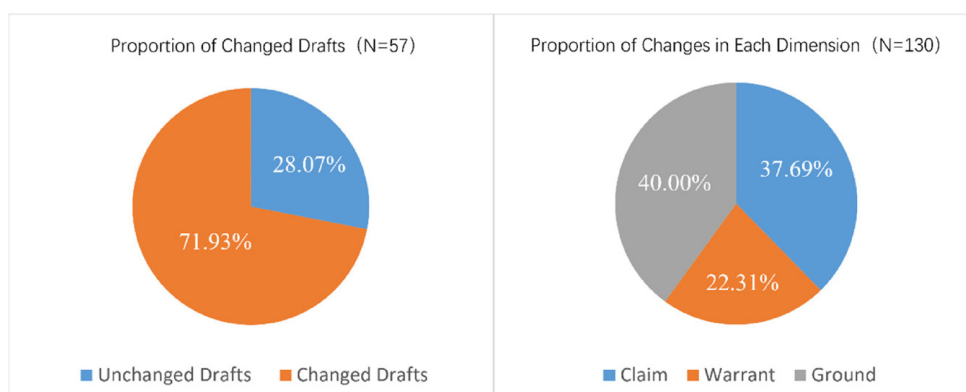


Figure 1. The proportion of improved drafts.

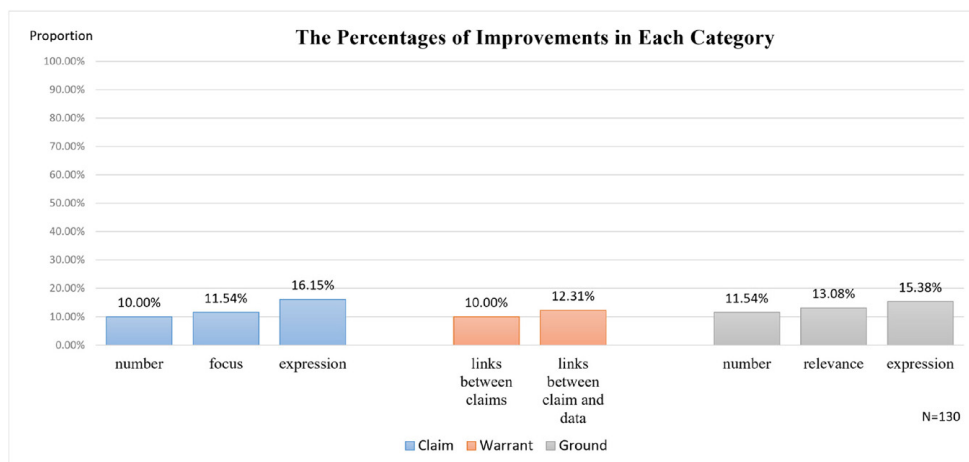


Figure 2. The percentage of improvements in each category.

15 (11.54%) narrowed the focus of their claims, and 21 (16.15%) improved the expression of their claims. Within the warrant category, 13 out of 130 (10%) modifications strengthened the connections among the claims, whereas 16 (12.31%) enhanced the links between the claims and the supporting data. In the ground category, 15 out of 130 (11.54%) modifications adjusted the number of grounds, 17 (13.08%) refined the relevance of their grounds, and 20 (15.38%) improved the linguistic expression of the grounds.

The results indicate that the participants demonstrated growing critical thinking skills applied to their writing, with the majority (71.93%) either optimizing the argument's structure, strengthening the links between points and evidence, or improving the clarity of arguments in the revised drafts.

Thematic Analysis of AI–Students Conversations

The conversations between the AI chatbot and students were systematically evaluated using a thematic analysis approach (Braun & Clarke, 2006; Clarke & Braun, 2017). Examples of conversations affiliated with to each theme are demonstrated in Figure 3.

Initially, open coding was applied to identify the fundamental concepts within each conversation. Next, axial coding was utilized to group related codes together; Finally, selective coding was used to synthesize three major themes, as described below.

- A. *Giving suggestions for revising the language:* Students consistently received feedback from the AI chatbot regarding grammatical accuracy, syntactical fluency, and clarity of expression.

A. Giving Language Revision Suggestions		
Sample Conversation	How to speak more specific of this sentence? No matter teachers or students, people in hold strict attitude toward language learning. Everyone at our university, whether teachers or students, has a serious approach to language learning. Please help me speak it more advanced. _____ is located in the mountains and has many trains in campus.	Can you tell me a more authentic expression of beautiful surroundings in and around the college? 关于“大学周围美丽的环境”，这里有一些更地道的表达方式： Stunning natural beauty surrounding the college Breathtaking scenery on the college campus and its surroundings Exquisite natural landscape surrounding the institution Picturesque environment encompassing the college campus Serene and captivating scenery around the college
B. Extending and Development Main Points		
Sample Conversation	How can I describe _____ University as an important training base for foreign language talent? University is an excellent institution that provides excellent training and education for foreign language talents. It has a strong faculty, excellent teaching facilities, and a wide range of international exchange programs. It cultivates foreign language talents who are proficient in language skills, have a broad international outlook, and are well-prepared for careers in various fields.	If my speech's central idea is: Using TalkwithAI App in English speaking class has great benefits. How to write my main points? Here are your main points: 1. TalkwithAI App can provide authentic materials for English speaking class, which helps students understand English in a real-world environment.
C. Providing Supporting Materials for Arguments		
Sample Conversation	Can you give me some examples to prove that observe the world without a smartphone could help us discover new beauty and fun stuff? Examples: 1. Noticing a beautiful sunset or a rainbow without the distraction of a smartphone screen. 2. Having a face-to-face conversation with a friend or family member, instead of texting. 3. Exploring a new area or park without relying on GPS. 4. Discovering a hidden gem of a restaurant or shop while walking down the street. 5. Observing nature and wildlife up close, without the barrier of a screen.	Please offer me some specific example to prove that staring at smartphone impairs our sight and easy to get a cataract when we get old. Examples: 1. Studies show that staring at a smartphone's screen for long periods can strain the eyes, leading to blurred vision or eye fatigue. 2. Prolonged screen exposure can increase the risk of developing cataracts or macular degeneration later in life.

Figure 3. Examples of conversations between the AI chatbot and students.

- B. *Extending and developing main points*: This theme revealed that students and the chatbot were observed to expand upon each other's ideas, resulting in a more profound exploration of the topics discussed.
- C. *Providing supporting materials for arguments*: The chatbot frequently enriched discussions by providing relevant materials, such as articles, research findings, and statistical data to strengthen the students' arguments.

Discussion and Conclusion

In the discussion of our study, we highlight the linguistic improvements observed in participants' speech script writing, particularly with a notable enhancement in lexical complexity. The participants' writing also reflected an escalation in critical thinking skills, as evidenced by the majority's refined argument structures, robust linkages between points and evidence, and heightened expression of arguments within the revised drafts. Our analysis of the conversations between the AI chatbot and students revealed three pivotal roles of the AI chatbot: refining linguistic expression, expanding ideas and perspectives, and supplying supporting materials. These insights implied a correlation between the improvements in the students' writing and the support by the AI chatbot. However, it is important to acknowledge that the development of critical thinking skills was not uniform among all participants; 28% failed to exhibit progress in this domain. This stagnation might be linked to the lack of targeted guidance on leveraging the AI chatbot effectively. Moving forward, educators are encouraged to devise more precise and pedagogically aligned conversational strategies with AI, customized to meet specific educational goals, content, and students' usage habits. This study also calls for increased attention from future researchers on how L2 writing instruction can cultivate learners' critical thinking, creativity, and other higher-order cognitive skills during the transition of foreign language education into the AI era.

References

- Barrot, J. S. (2023). Using ChatGPT for second language writing: pitfalls and potentials. *Assessing Writing*, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- Clarke, V., & Braun, V. (2017). Thematic analysis. *The Journal of Positive Psychology*, 12(3), 297–298. <https://doi.org/10.1080/17439760.2016.1262613>

- Cobb, T. (2002). VocabProfilers. <http://www.lex tutor.ca/vp/>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing*, 54, 100666. <https://doi.org/10.1016/j.asw.2022.100666>
- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE and FREQUENCY programs. <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>
- Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: doing critical participatory action research*. Springer.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474–496. <https://doi.org/10.1075/ijcl.15.4.02lu>
- Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, 96(2), 190–208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x
- OpenAI. (2022). Introducing ChatGPT. Open AI. <https://openai.com/blog/chatgpt>
- 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>
- Tan, L. (2024). NeoSCA (version 0.1.1) [Computer software]. <https://github.com/tanloong/neosca>
- Thorp, H. H. (2023). ChatGPT is fun, but not an author. *Science*, 379(6630), 313–313. <https://doi.org/10.7759/cureus.37306>
- Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.
- Wen, Q. F. (2017). The production-oriented approach: a pedagogical innovation in university English teaching in China. In L. J. Z. Wong and K. Hyland (Eds.), *Faces of English: students, teachers, and pedagogy* (pp. 91–106). Routledge.
- Wen, Q. F. (2018). The production-oriented approach to teaching university students English in China. *Language Teaching*, 51(4), 526–540. <https://doi.org/10.1017/S026144481600001X>
- Zhang, M., & Li, J. (2021). A commentary of GPT-3 in *MIT Technology Review 2021. Fundamental Research*, 1(6), 831–833. <https://doi.org/10.1016/j.fmre.2021.11.011>
- Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: evidence from English narrative writing. *System*, 118, 103141. <https://doi.org/10.1016/j.system.2023.103141>