

Cognitive Engagement of Low Proficiency Learners: Evaluating the Impact of ChatGPT Feedback in EFL Writing

Yingxia Li^{1,a,*}, Wanyi Du^{1,b}, Bike Zhu^{1,c}

¹School of Foreign Languages, Dalian University of Technology, Dalian, China

^aliyixia@dlut.edu.cn, ^bduwanyi@dlut.edu.cn, ^czbkangela0617@dlut.edu.cn

*Corresponding author

Abstract: ChatGPT, a large-scale language model, holds significant potential for providing consistent and timely feedback, thereby offering comprehensive support across various dimensions of writing. However, existing research is limited in its exploration of the specific impact of ChatGPT's written feedback on English as a Foreign Language (EFL) learners with low language proficiency. This study delves into the effects of ChatGPT feedback on such learners, utilizing data from prompt-feedback interactions and subsequent interview analyses. The findings reveal that the essay reviewing process, involving interactions with ChatGPT in English, posed substantial challenges for the learners. Additionally, the study suggests that ChatGPT's written feedback has limitations in effectively enhancing the writing skills of learners in this proficiency category.

Keywords: ChatGPT, Written Feedback, Low Proficiency Learners

1. Introduction

The emergence of ChatGPT has heralded a revolutionary transformation in the realm of second language teaching, resulting in a surge of scholarly literature in aiding English language teaching and learner development. The majority of research findings indicate substantial enhancements in academic performance when aided by AI technologies [1]. Hsieh et al. (2020) found that the presence of robots positively impacted student confidence in the learning process. Furthermore, the use of AI chatbots and tutors, as highlighted by Yang & Shulruf (2019), not only boosted students' confidence but also contributed to a reduction in learning anxiety. Despite two studies reporting no statistically significant change in academic performance, both Topal et al. (2021) and Yang & Shulruf (2019) observed increased learning interest and confidence among students. While AI can be an invaluable tool in the assessment and feedback process, it is not without its limitations, particularly in comprehending intricate subject matter, contextual nuances, and the distinctive qualities of individual students' work [4][5]. Moreover, there is a lack of standardized criteria for evaluating the appropriateness of grading systems across diverse contexts [6]. More importantly, the intelligent system primarily contributes in delivering feedback that enables learners to engage in meaningful discussions with GPT, following the receipt of feedback. It is the dynamic process of interactions empower learners and augment their writing abilities. However, studies on the personalized learning offered by AI technologies still needs more evidence [7]. To fill in this gap, this study takes a qualitative approach to examine the cognitive engagement of low proficiency (LP) in the writing process with the chatbot.

2. Literature review

Research on taking advantages of intelligent tools to provide writing feedback to EFL learners dates back to AWE systems [8], such as *Criterion*, *WriteToLearn*, *PiGai*, etc. [9]. Research report that users of *Grammarly* demonstrate superior performance compared to their counterparts in second language (L2) writing tasks [10], and the positive impact of AI-based written feedback on students' motivation and self-efficacy for L2 writing [11]. Initial attempts to integrate a chatbot-centered approach for supporting second language (L2) writing among learners also yielded commendable results [12].

Unlike Grammarly, ChatGPT goes beyond grammar and style suggestions, providing assistance with structure, coherence, and argumentation. It excels in understanding and responding to the broader context

of text, offering more nuanced feedback. Additionally, ChatGPT can generate creative content suggestions and assist with idea generation, making it a valuable tool for writers seeking a more diverse and holistic writing assistance. Studies on the impact of ChatGPT in the field of education cover a range of dimensions, such as enhancing creativity and critical thinking [13], aiding in medical exam preparation [14], and empowering learners in engineering education [15]. Texts generated by ChatGPT received acclaim for their overall quality, original content, and proficiency in composing literature reviews [16].

While the above-mentioned studies have explored the assistive role of AI in written feedbacks, few studies concern with how much ChatGPT empowers LP (low proficiency) learners' writing abilities, and evidence on the relationship between the effectiveness of prompts and that of ChatGPT's feedback is scarce. To narrow the research gap, the study has used ChatGPT 3.5 to investigate how eight LP students engaged with ChatGPT to revise their writings based on the feedback provided. This qualitative approach is guided by the following research questions:

- (1) What are the general features of feedbacks generated by proficiency learners?
- (2) How do LP learners cognitively engage in processing feedback provided by ChatGPT?

3. Methods

3.1 Context and Participants

All together eight undergraduates aged 19 to 20 participated voluntarily. They are taught by the same teacher for two semesters. All the students are majoring in software engineering and have attended one semester of the English writing course. Both their writing scores and CET 4 (a national college English test in China) scores are below the average in the whole grade.

3.2 Data Collection

Data sources included student texts, student screenshots of conversations (prompts and responses) with ChatGPT, and student interviews.

In stage 1, all the participants were given an essay topic "Is technology making people less creative?". They were expected to work out an essay with 200 to 300 words and under the framework required by the writing course. They were required to submit their first E-copy draft within a week. During the process, they were not allowed to use any intelligent tools to aid their writing. In stage 2, trainings were provided to the participants on how to create effective prompts in English. Participants were instructed to give clear instructions by providing a specific request (e.g., grammar check, sentence structure advice, vocabulary suggestions), suggest a clear reference point (e.g., "Please review the second paragraph"), and review suggestions to decide whether to accept or reject suggestions. Emphasis was put on understanding the reasons behind the feedback rather than getting corrections. In stage 3, participants interacted with ChatGPT to elicit feedbacks on their essays in English. They had one week to revise their essay and were required to save the screenshots of the whole process of interactions with ChatGPT. In stage 4, participants submitted their screenshots and final-drafted essay. In stage 5, each student attended a one-on-one interview (in Mandarin) around 30 min. All the interviews were recorded and saved in audio track.

3.3 Data Analysis

The quantity and quality of ChatGPT feedbacks were calculated and analyzed. Participants were required to count the number of the effective and ineffective feedbacks produced by ChatGPT. They were supposed to analyze the ineffective feedback by providing the causes of failure. Additionally, the participants' uptake of GPT's feedback is examined, considering three dimensions: content, organization, and language. The quantity of uptakes in each of these three categories is calculated accordingly.

An additional coder Teacher L was invited to join the research to code the interview transcriptions. She graduated from a US university with a doctor degree in pedagogy and has more than 30 years of teaching experience. All the coding were done by L. The researcher proofread the auto transcriptions from Tencent Meeting, a software application in China which allows users to hold meetings online. It can automatically transcribe the audio into written words.

4. Results

4.1 General patterns in ChatGPT WFC

4.1.1 The quantity and perceived quality of feedbacks

Of the number of feedbacks, the highest reached 24, while the lowest was only 12. Of the total feedbacks produced, 80.27% are effective, 17% are ineffective, and 2.04% are categorized as “unknown”, as reported by the participants.

4.1.2 Ratings of ChatGPT WFC

Table 1 presents the rating results of quality and helpfulness of the feedbacks provided by ChatGPT. Five participants rated 5 for all three aspects of the quality and helpfulness, indicating they were generally satisfied with the whole writing practice. Student W thought that ChatGPT could point out the problems of his essay, but he admitted that sometimes he could not fully understand the feedback. He had to turn to other machine translation for help, because using Chinese was not allowed by the research. Student Y noted that she enjoyed the writing process with ChatGPT, but sometimes she didn't know how to handle with the feedback. She expressed her concern, “The feedback is good, but I still don't know what to do even if I understand where my problems are.” When asked for an example, she said, “For example, ChatGPT said I need more examples to support my claim, but the problem is I don't know what examples I should write.”

However, three participants (L, Z, K) give high ratings to the quality of ChatGPT's feedback yet rated low for the helpfulness of supporting writing category. Student L reported in the subsequent interview that he thought the feedback was good, but he had hard time to figure out the meaning. He said that it would be much better if Chinese was allowed during the interaction process. Student Z and K also expressed the same request. Student Z said, “My English is not good, so it's hard for me to use English with ChatGPT. Many times, I spent a lot of efforts in thinking about how to put the prompts in English before I finally get the feedback. It was hard, though I understand this helps me to think in English.” Student K expressed more directly about the wish to use Chinese instead of English when interacting with ChatGPT. “The amazing part of ChatGPT is you can use whatever language you want, and it can understand you. That's why I think the language should not be limited to the use of English. That's what technology for.”

Most participants (Student W, S, P, L, Zh) mentioned about reduced anxiety when interacting with ChatGPT compared with feedback provided by humans, either teacher feedback or peer feedback. “I like chatting with machines more than humans. I feel less stressed. If I don't understand something, I can ask GPT to translate for me...”, Student S said.

Student Zh also mentioned the concerns of plagiarism. “I asked ChatGPT to write a paragraph for me, and it did! I think it's perfect. I don't have to change anything, but can I just copy and paste?”

Table 1: Scores on students' perceived ChatGPT feedback

Students	Perceived quality of ChatGPT feedback			Perceived helpfulness for improving writings
	Content	Organization	Language	
Y	5	5	5	5
W	5	5	5	5
L	5	4	5	4
Z	4	5	5	3
S	5	5	5	5
Zh	5	5	5	5
P	5	5	5	5
K	5	5	5	3

5. Discussion

5.1 Feedback and language proficiency

ChatGPT as a large-scaled language model can provide consistent and timely feedback, handle large volumes of work, and address specific language-related issues effectively. The biggest difference

between ChatGPT feedback and human feedback is the quality of ChatGPT WFC depends more on the quality of prompts. The effectiveness and accuracy of prompts determine the accuracy and effectiveness of the feedback. Advance learners with high proficiency in English may have high expectations for nuanced and contextually relevant feedback. They might find ChatGPT's suggestions too generic for their advanced writing needs^[17]. For low proficiency learners, on the other hand, they face much more challenges when interacting with ChatGPT, especially in English. When they fail to produce clear and accurate prompts, the quality of feedback is reduced accordingly. Prompts trainings are provided before the essay reviewing process. However, some participants still use prompts like “review my essay” or “comment on my essay”, resulting in lengthy and complicated feedback. When this happens, these participants often feel helpless and lost. In one case, the participant regarded the reviewing process too hard, as he needed to constantly produce new prompts in English. “I didn’t expect it would be so hard,” he reported, “I had thought using ChatGPT would have made my work easier.”

5.2 Cognitive engagement of LP learners

5.2.1 Difficulty in Articulating Questions

Low-proficiency learners often struggle to articulate questions or concerns clearly, posing challenges in seeking precise assistance from ChatGPT. As this research categorizes learners by their English level, participants are mandated to use English throughout their interactions with ChatGPT. Despite pre-experiment training provided by the researcher, participants still find generating prompts challenging. In contrast, human feedback, including teacher and peer feedback, doesn’t necessitate prompts. The feedback is structured based on specific criteria, making it more targeted. For LP learners, human feedback is friendlier, albeit less accessible compared to ChatGPT or other AWE.

5.2.2 Overreliance on AI

Low-proficiency learners may overly rely on AI feedback, neglecting the development of their writing and revision skills. The reviewing process presents significant challenges for these students. Instead of seeking written feedback and revising their work, LP students find it easier to directly request ChatGPT to write a paragraph for them. This inclination stems from a perceived poor English level and a fear mentality associated with using English. Additionally, they recognize that writing involves not only words and sentences but also a particular way of thinking. Consequently, when ChatGPT provides feedback, learners are more inclined to accept it without reservation.

5.2.3 Limited Improvement in Writing Proficiency

ChatGPT can provide immediate corrections, but its contribution to long-term improvements in writing proficiency is limited. Substantial progress in writing often requires extended periods of comprehensive feedback, guidance, and practice. While AI tools, like ChatGPT, offer valuable assistance, they cannot replace the sustained effort and guidance necessary for lasting improvement in writing skills.

Despite AI, such as ChatGPT, being a valuable tool in assessment and feedback, it has limitations in understanding complex content, context, and the unique qualities of individual student work. In contrast, human teachers are better equipped to provide expertise, contextual understanding, personalized guidance, and emotional support—qualities that AI, including ChatGPT, cannot fully replicate.

References

- [1] Kahn, K., & Winters, N. (2021). *Constructionism and AI: A history and possible futures*. *British Journal of Educational Technology*, 52(3), 1130–1142. <https://doi.org/10.1111/bjet.13088>
- [2] Hsieh, Y. Z., Lin, S. S., Luo, Y. C., Jeng, Y. L., Tan, S. W., Chen, C. R., & Chiang, P. Y. (2020). *ARCS-assisted teaching robots based on anticipatory computing and emotional Big Data for improving sustainable learning efficiency and motivation*. *Sustainability (Switzerland)*, 12(14). <https://doi.org/10.3390/su12145605>.
- [3] Yang, Y. Y., & Shulruf, B. (2019). *Expert-led and artificial intelligence (AI) system-assisted tutoring course increase confidence of Chinese medical interns on suturing and ligature skills: Prospective pilot study*. *Journal of Educational Evaluation for Health Professions*, 16. <https://doi.org/10.3352/jeehp.2019.16.7>
- [4] Topal, A. D., Eren, C. D., & Gecer, A. K. (2021). *Chatbot application in a 5th grade science course*. *Education and Information Technologies*, 26(5), 6241–6265. <https://doi.org/10.1007/s10639-021-10627-8>

- [5] Ma, H., & Slater, T. (2015). Using the developmental path of cause to bridge the gap between AWE scores and writing teachers' evaluations. *Writing & Pedagogy*, 7(2-3), 395-422. <https://doi.org/10.1558/wap.v7i2-3.26376>
- [6] Sun, Y. (2021). Application of artificial intelligence in the cultivation of art design professionals. *International Journal of Emerging Technologies in Learning*, 16(8), 221-237. <https://doi.org/10.3991/ijet.v16i08.22131>
- [7] Chew, E., & Chua, X. N. (2020). Robotic Chinese language tutor: Personalising progress assessment and feedback or taking over your job? *On the Horizon*, 28(3), 113-124. <https://doi.org/10.1108/OTH-04-2020-0015>
- [8] Dikli, S. (2010). The nature of automated essay scoring feedback. *CALICO Journal*, 28, 99-134.
- [9] Weigle, S. C. (2013). English as a second language writing and automated essay evaluation. in M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 36-54). new york, ny: Routledge.
- [10] Dizon, G., & Gayed, J. M. (2021). Examining the Impact of Grammarly on the Quality of Mobile L2 Writing. *JALT CALL Journal*, 17(2), 74-92. <https://doi.org/10.29140/jaltcall.v17n2.336>
- [11] Nazari, N., Shabbir, M. S., & Setiawan, R. (2021). Application of Artificial Intelligence powered digital writing assistant in higher education: Randomized controlled trial. *Heliyon*, 7(5), e07014. <https://doi.org/10.1016/j.heliyon.2021.e07014>
- [12] 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>
- [13] Zhai, X. (2022). ChatGPT user experience: Implications for education. Available at SSRN4312418.
- [14] Gilson, A., Safranek, C., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2022). How well does ChatGPT do when taking the medical licensing exams? The implications of large language models for medical education and knowledge assessment. *medRxiv*, 1-9. <https://doi.org/10.1101/2022.12.23.22283901>
- [15] Qadir, J. (2022). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. *TechRxiv. Preprint*. <https://doi.org/10.36227/techrxiv.21789434.v1>
- [16] Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *BioRxiv*, 2022-12.
- [17] Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 1-29. <https://doi.org/10.1007/s10639-023-12146-0>