

ChatGPT-Assisted English Writing Development Among Higher Education Students

Dr. Nattapong Srisuk^{1*} and Dr. Siriporn Chaiyaporn²

^{1*}Department of English, Thammasat University, Bangkok, Thailand.

²Department of English, Thammasat University, Bangkok, Thailand.

Abstract: The swift incorporation of Artificial Intelligence (AI) systems in education is providing new possibilities for enhancing second and foreign language writing skills. ChatGPT is one of the tools that has become a readily available AI that can offer very quick feedback, suggestions, and amendments for students' writing. In spite of the interest in this issue, there is still not enough information on how ChatGPT affects students' academic performance, self-efficacy, and revision strategies. The present paper aims to analyze how effective ChatGPT-based feedback can be for the development of students' written English. To achieve this goal, the researchers used a mixed-methods approach to research the issue. A total of 41 students participated in the study. The writing samples underwent thorough analysis for accuracy, coherence, and lexical richness of students' written work after six weeks of ChatGPT-assisted intervention. The results show that students' writing has significantly improved in terms of correctness and structural coherence. At the same time, students expressed quite positive opinions about ChatGPT but claimed that the use of such an AI could lead to dependency and a lowering of critical thinking ability. The findings add new evidence to the existing literature on automated assessment of writing and AI language learning and can be useful for professionals working in the area of curriculum design. The researchers made recommendations on the modern use of ChatGPT in education.

Key Words: chatgpt, AI-assisted WRITING, english language learning, higher education, writing feedback

(Received: 16 December 2022; Revised: 31 January 2023; Accepted: 20 February 2023; Published: 31 March 2023)

Introduction

Writing stands out as one of the most challenging language skills for English as a Foreign/ Second Language (EFL/ESL) learner in their higher education experiences, as it demands the simultaneous control of grammar, vocabulary, coherence, and rhetorical conventions at the same time (Ali, 2020). Traditional feedback mechanisms rely heavily on the availability of instructors for the feedback process and lead to ineffective support for learners in the form of timely corrective input on their writing skills (Kessler, 2020). With the advent of the last decade, Artificial Intelligence has become an ever-growing solution that can help ease these barriers in the process, as automated writing evaluation (AWE) systems and AI writing assistants have the potential to deliver scalable on-demand feedback (Nazari et al., 2021).

While tools like Grammarly have proven beneficial to learners with grammar accuracy and with getting access to corrective feedback, it has been shown that automated written corrective feedback is able to influence research quality and its interaction with metalinguistic information of different types as well (Guo et al., 2022; Huang et al., 2020; Jang, 2020). Besides the rule-based correction tools, the AI-driven Pigai tool has also been successfully used to provide formative feedback in EFL settings, with long-term studies suggesting that the AWE feedback can lead to long-lasting positive changes in the writing skills of the learners (Jingxin & Razali, 2020; Lee, 2020).

The birth of generative agents based on large language models like ChatGPT means moving away from rule-based correction and using generative dialogic feedback, which has a more interactive nature that can explain, rewrite, and brainstorm ideas together with the learner (Divekar et al., 2022). There are multiple implications of such a shift for the efficiency of such tools and the question of writing self-efficacy, technological skills, and other aspects of academic and digital communication take place (Kessler, 2020; Lee & Evans, 2019). Even though AWE studies serve as great background for understanding the technology, the details of the conversational nature of ChatGPT are still not clear, and you need to conduct this study in order to learn more about it.

The significance of this study is to contribute to the ongoing change in the area of generative AI for language learning through the examination of the example of ChatGPT and its effect on the academic writing skills of the students of higher education institutions, taking it as the improvement of the earlier tools like Grammarly and Pigai. The study makes use of the discussion of pre- and post-writing skills by applying both quantitative and qualitative methods of analysis, and this helps to increase the knowledge about how the learners perceive the feedback of conversational AI. Moreover, the guidance for instructors on the creation of adequate AI writing tasks is offered.

The rest of the paper is organized in a logical way, as Section II will present the earlier studies on AWE, AI-based feedback, and conversational agents in language learning, ending with the reason for motivating the current research questions. Section III focuses on the research design, participants and procedures used in this study and present the flowchart for providing visual information about the process. Section IV presents all the findings and discusses them organized in three subsections related to writing skills, student perception and sub-skills of students' writing. Section V contains the discussion of how the findings can be related to the field in general, which will lead us to the final part of the paper containing implications, limitations and suggestions for the future research.

Mapping the Terrain: Know About AI-Assisted Writing Feedback

The existing literature indicates that AI-enabled evaluation would improve particular aspects of learner writing. Automation enabled writing evaluation to have an impact on students' revising behavior which may also influence the nature of teachers' feedback since automated evaluation and human feedback interact with each other (Link et al., 2022). It has been also found that students' digital competence serves as a mediating variable in effective students' use of automated evaluation tools since digital skills are not equal among university students in different European countries (López-Meneses et al., 2020). RCT data proved that an AI-based digital writing assistant significantly improved writing results of students at universities compared to the control group (Nazari et al., 2021). In addition, studies show that students have a generally positive opinion on the idea of using automated evaluation although the effectiveness of the method is doubted in terms of some higher order issues related to the continuity of argumentation and coherence (ONEILL & Russell, 2019). In general, it was found that not only standardized writing assessment appeared to improve students' writing, but one has to be careful not to consider it as an all-in-one method of teaching writing Parra G. & Calero S., (2019). The research dedicated to improving student engagement has focused on the interaction of students with AI feedback in terms of behavioral activity rather than the improvement of scores. The students have shown varied levels of cognitive, behavioral, and effective engagement with automated feedback technology, while the collaborative technology-enhanced writing environment promotes the quality of writing performance and self-regulation (Koltovskaia, 2020; Rahimi & Fathi, 2022). Even though automating the feedback has been proved to produce results, language difficulties still persist among EFL learners, which means that not all mistakes that can be corrected through AWE technology (Tambunan et al., 2022). In addition to text-based corrections, conversational and

embodied AI has been examined in relation to chatbots assisting with punctuation instruction and helping kids practice reading (Vázquez-Cano et al., 2021; Xu et al., 2021). There are systematic reviews proving that chatbot application in the language learning environment yields positive, yet methodological results (Huang et al., 2022). Findings from perception studies on automated feedback taken across different student cohorts and synoptic reviews of AI in education as a three-paradigmatic field prove that it is important to implement technology properly together with a suitable pedagogical approach (Miranty & Widiati, 2021; Ouyang & Jiao, 2021).

Where the gap is. In conclusion, the available studies show similar tendency: automatic/AW feedback effectively improves learning processes in terms of writing technicalities and language mechanics, but its impact on higher cognitive skills is yet to be studied. Unfortunately, the majority of existing studies cover only rule-based (strict-to-function) AWE technologies developed before the advent of large language generative models. This calls into question whether the latest generation dialogic technology known as ChatGPT leads to different results in terms of student engagement and outcome and, thus, learner independence, critical thinking, etc.

Behind the Study: Design, Participants, and Procedure

Conducting a Two-Phase Study

This research utilized a methodology that is mixed in nature with a quasi-experimental design that integrates a pre-test/post-test writing assessment within the same group with a follow up cross-sectional perception survey. The approach allowed for the determination of the variation in the performance due to the implementation of ChatGPT in the writing process, while qualitative data were collected from students regarding their opinions about AI use for writing and their experiences in the process. The data-gathering structure was necessary to eliminate the shortcomings of previous AWE efforts when the improvement is clearly visible, however, no data about the engagement of students is provided. The intervention took place within six weeks, and it was incorporated into a standard English for Academic Purposes course that makes research ecologically valid.

Participants of the Study

The participants were students from an EAP course who were selected according to the purposive sampling technique, which allowed to ensure the same level of exposure to the learning environment as well as technological awareness of academic writing tools. It is important to mention that non-native speakers of English were included in analysis, without any prior knowledge of AI-assisted writing approaches, with the aim to single out the impact of the intervention in question. The students were aware of the aim of the research and signed the informed consent forms guaranteeing that their participation and performance will not affect their current standing in the course. Demographic details were gained in order to provide information about the sample.

Instruments and Procedures Used for Data Collection and Analysis

The research utilizes three tools for data collection that included pre-intervention essay, post-intervention essay produced in the same conditions after the application of ChatGPT writing process, and perception survey conducted after the intervention. The essays were analyzed on the basis of an analytical rubric evaluating all components of writing task, among which were grammatical accuracy, lexical range, coherence and cohesion, and task completion. The percentages of correct answers were

calculated, with the help of statistics for the purpose of determining the effectiveness of the intervention. The overall data collection and analysis procedure is summarized in figure 1.

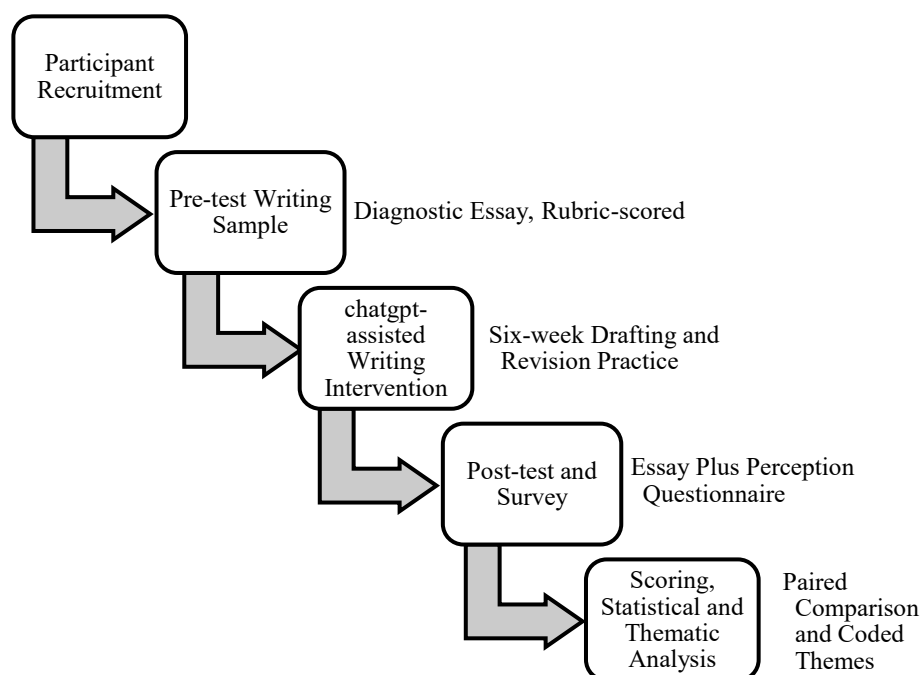


Figure 1: Experimental Workflow for Evaluating ChatGPT-Assisted Writing in English Education

Figure 1 illustrates the sequential flow of the study: from participant recruitment and baseline (pre-test) writing assessment, through the six-week ChatGPT-assisted writing intervention, to the post-test assessment and perception survey, concluding with rubric scoring, statistical comparison, and thematic analysis.

Signals from the Classroom

Writing Performance Through Pre- and Post-Testing

Statistical comparisons of pre- and post-tests indicated a significant increase in scores in all four rubric suggestions. While the biggest improvement was recorded in grammar volume, lexical areas were in the second place, while coherence and cohesion and task achievement were improved much less. These facts indicate that it worked best for superficial and lexical aspects, with improvement in more complicated organizational qualities taking a bit more effort on the teacher's part. Figure 2 shows average scores from rubric sections before and after the experiment.

Perception and Engagement of Students

Preliminary analysis of the perception questionnaire showed that students generally see ChatGPT as a useful tool for writing. This is particularly obvious in the answers connected to idea generation and rapid error correction. However, the levels of agreement are lower for the answers about trust in ChatGPT suggestions on argument development and perception of the possibility to use ChatGPT alone without consulting the instructor. One of the most popular themes in the given responses was the acknowledgment of AI feedback being immediate and unemotional but a doubt represented by the possibility of accepting changes produced by the software without knowing the reasoning behind it. Figure 3 shows the average estimates for agreement in regard to five statements included in the survey.

Analysis on the Level of Subskills

The analysis of written papers by subskills has shown that not all errors were decreasing persistently. It may be seen that mechanical and grammatical mistakes have reduced the most compared to the previous analysis. The case with mistakes associated with coherence and cohesion is less optimistic. Issues related to argumentation and critical point of view were reduced the least. This explains the correlation of the students' trust in ChatGPT being higher for mechanical mistakes and lower for serious argumentation issues.

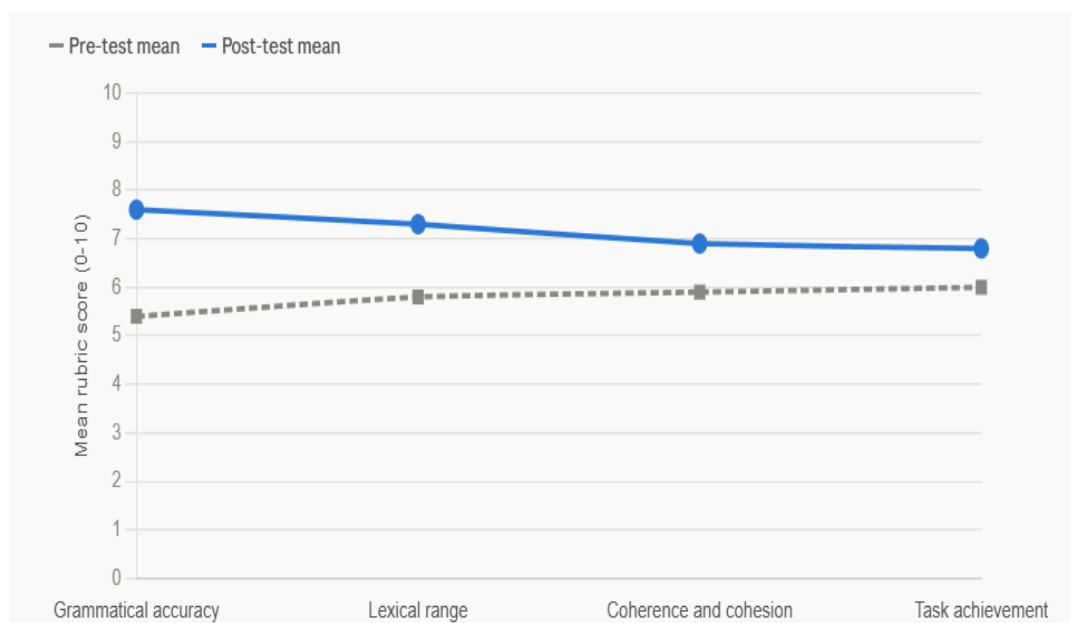


Figure 2: Comparison of Pre-Test and Post-Test Mean Writing Scores Across Assessment Criteria

Figure 2 shows Line chart comparing mean pre-test (dashed gray, square markers) and post-test (solid blue, round markers) rubric scores (0–10 scale) across four writing dimensions. The gap between lines is widest for grammatical accuracy and narrowest for task achievement.

The Trust Gradient in AI-Assisted Writing

The findings of the study point to the conclusion that ChatGPT can bring the best results when handling writing issues that are simple and clear-cut, while not improving such matters as coherence or argumentation significantly. This trend reflects a common case in writing with AI technologies. Tools that are perfect in correcting syntax do not help students in learning to compose their texts. Moreover, the fact that students are less confident using ChatGPT independently than when using certain functions of the tool shows that students recognize what the technology is good at and what kind of tasks it cannot perform. The results of the research underline the necessity for introducing ChatGPT into educational curriculums properly. Rather than using it only as a means of correction of texts, teachers should help students understand how to assess the suggestions of AI. The findings also indicate that considerable success in the area of mechanical accuracy can be achieved pretty soon, whilst developing an ability to use the AI feedback wisely takes longer.

Conclusion

The impact of ChatGPT-assisted eight weeks writing intervention on the writing skill and attitude of university students were analyzed in this study. According to the results of the research, students achieved

significant improvement in grammatical correctness and vocabulary range used and minor improvements in coherence and writing achievement. Moreover, students' opinions towards ChatGPT were mostly positive with their comments especially on its effectiveness in idea generation and fast error correction. However, students had a lower feeling of confidence when using the tool as an independent resource especially in writing arguments. The current findings refer to the previously conducted studies on the automated writing evaluation wherein conversational format; generative AI technology usage demonstrates its feedback profile that is directed towards explaining its technical effectiveness and giving insight into the issue of student-centered approach combined with critical thinking in interpreting computer-generated messages. However, the shortcomings of the conducted research are represented by its short period of implementation, limited number of respondents, and the means of collecting information through perceptions of the participants instead of rubric-based assessments. Further investigations in the field will require recruiting a bigger group of respondents, as well as designing a study that would compare ChatGPT with the other AI writing tools.

References

- Ali, Z. (2020, February). Artificial intelligence (AI): A review of its uses in language teaching and learning. In *IOP Conference Series: Materials Science and Engineering* (Vol. 769, No. 1, p. 012043). IOP Publishing. <https://doi.org/1088/1757-899X/769/1/012043>
- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y., Su, H., Chen, Y., ... & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: design and evaluation. *Computer Assisted Language Learning*, 35(9), 2332-2360. <https://doi.org/10.1080/09588221.2021.1879162>
- Guo, Q., Feng, R., & Hua, Y. (2022). How effectively can EFL students use automated written corrective feedback (AWCF) in research writing? *Computer Assisted Language Learning*, 35(9), 2312-2331. <https://doi.org/10.1080/09588221.2021.1879161>
- Huang, H. W., Li, Z., & Taylor, L. (2020, May). The effectiveness of using Grammarly to improve students' writing skills. In *Proceedings of the 5th international conference on distance education and learning* (pp. 122-127). <https://doi.org/10.1145/3402569.3402594>
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of computer assisted learning*, 38(1), 237-257. <https://doi.org/10.1111/jcal.12610>
- Jang, S. S. (2020). The Efficacy of Different Types of Metalinguistic Information in L2 Written Corrective Feedback. *English teaching*, 75(4), 33-56. <https://doi.org/10.15858/engtea.75.4.202012.33>
- Jingxin, G., & Razali, A. B. (2020). Tapping the potential of Pigai automated writing evaluation (AWE) program to give feedback on EFL writing. *Universal Journal of Educational Research*, 8(12B), 8334-8343. <https://doi.org/10.13189/ujer.2020.082638>
- Kessler, M. (2020). Technology-mediated writing: Exploring incoming graduate students' L2 writing strategies with Activity Theory. *Computers and Composition*, 55, 102542. <https://doi.org/10.1016/j.compcom.2020.102542>
- Koltovskaia, S. (2020). Student engagement with automated written corrective feedback (AWCF) provided by Grammarly: A multiple case study. *Assessing Writing*, 44, 100450. <https://doi.org/10.1016/j.asw.2020.100450>
- Lee, M. K., & Evans, M. (2019). Investigating the operating mechanisms of the sources of L2 writing self-efficacy at the stages of giving and receiving peer feedback. *The Modern Language Journal*, 103(4), 831-847. <https://doi.org/10.1111/modl.12598>

- Lee, Y. J. (2020). The long-term effect of automated writing evaluation feedback on writing development. *English Teaching*, 75(1), 67-92. <https://doi.org/10.15858/engtea.75.1.202003.67>
- Link, S., Mehrzad, M., & Rahimi, M. (2022). Impact of automated writing evaluation on teacher feedback, student revision, and writing improvement. *Computer Assisted Language Learning*, 35(4), 605-634. <https://doi.org/10.1080/09588221.2020.1743323>
- López-Meneses, E., Sirignano, F. M., Vázquez-Cano, E., & Ramírez-Hurtado, J. M. (2020). University students' digital competence in three areas of the DigCom 2.1 model: A comparative study at three European universities. *Australasian Journal of Educational Technology*, 36(3), 69–88. <https://doi.org/10.14742/ajet.5583>
- Miranty, D., & Widiati, U. (2021). *Automated Writing Evaluation (AWE) in Higher Education: Indonesian EFL Students' Perceptions about Grammarly Use across Student Cohorts*. *Pegem Egitim ve Ogretim Dergisi*, 11 (4), 126–137. <https://doi.org/10.47750/pegegog.11.04.12>
- 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). <https://doi.org/10.1016/j.heliyon.2021.e07014>
- ONeill, R., & Russell, A. (2019). Stop! Grammar time: University students' perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, 35(1). <https://doi.org/10.14742/ajet.3795>
- Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, 2, 100020. <https://doi.org/10.1016/j.caeai.2021.100020>
- Parra G., L., & Calero S., X. (2019). Automated Writing Evaluation Tools in the Improvement of the Writing Skill. *International Journal of Instruction*, 12(2), 209-226. <https://doi.org/10.29333/iji.2019.12214a>
- Rahimi, M., & Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: a mixed methods study. *Computer Assisted Language Learning*, 35(9), 2627-2674. <https://doi.org/10.1080/09588221.2021.1888753>
- Tambunan, A. R. S., Andayani, W., Sari, W. S., & Lubis, F. K. (2022). Investigating EFL students' linguistic problems using Grammarly as automated writing evaluation feedback. *Indonesian Journal of Applied Linguistics*, 12(1), 16-27. <https://doi.org/10.17509/ijal.v12i1.46428>
- Vázquez-Cano, E., Mengual-Andrés, S., & López-Meneses, E. (2021). Chatbot to improve learning punctuation in Spanish and to enhance open and flexible learning environments. *International Journal of Educational Technology in Higher Education*, 18(1), 33. <https://doi.org/10.1186/s41239-021-00269-8>
- Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. *Computers & Education*, 161, 104059. <https://doi.org/10.1016/j.compedu.2020.104059>