

This work is licensed
under a Creative
Commons Attribution
4.0 International
License.

ChatGPT as a writing peer: Enhancing university students' writing skills through AI- assisted formative assessment

 **Bao Nguyen**

Hyogo University, Japan
ngdthbao@hyogo-dai.ac.jp

 **Jiale Lu**

Waseda University, Japan
jialelu@ruri.waseda.jp

This study explores how university students perceive and use ChatGPT as a tool to enhance their English academic writing skills under instructor guidance. Forty first-year Japanese students enrolled in an academic writing course at two private universities in Tokyo participated in the study. Over six weeks of structured training sessions that incorporated the use of ChatGPT, students engaged with the tool as a feedback partner to revise their writing drafts. A post-study questionnaire was administered to examine students' perceptions of ChatGPT's effectiveness and their overall reflections. The results indicated substantial improvement in scores across all four evaluation criteria (content, communicative achievement, organization, and language) in the second drafts after using ChatGPT. However, students also expressed some hesitation about integrating ChatGPT into their writing process and identified several challenges (for example, providing effective prompts to ChatGPT). The findings contribute to ongoing discussions on incorporating generative artificial intelligence (AI) into writing instruction, highlighting both the pedagogical potential and limitations of ChatGPT as a supportive peer tool in tertiary education.

Keywords: ChatGPT, academic writing, learner training, generative AI

Introduction and literature review

Artificial intelligence (AI) undoubtedly has revolutionized the field of education and become an indispensable tool for both educators and students. For teachers, by being able to capitalize on AI tools, not only can they enhance their practice, they can also benefit their students' learning journey. According to Kohnke et al.

(2023), the competencies to appropriately utilize those AI-tools that are based on Large Language Models (LLMs) may soon become necessary for language teachers to improve their teaching practice and enhance their career development. Current studies have investigated the efficacy of AI in face-to-face, blended, and online instruction (Kexin et al., 2020). Chaudhry and Kazim (2021) also argue that AI tools can also help teachers devise effective methodologies using learning data from students, resulting in effective assessment and feedback. While it goes without saying that the tool provides teachers with useful affordances for their teaching practice, a question that is worth contemplating is their effectiveness in assisting students' learning.

For students, studies have shown that thanks to their affordances, AI-tools, adapted appropriately in education, have the potential to enhance their competency. In particular, ChatGPT (short for Chat Generative Pre-trained Transformer) has been said to provide students with opportunities to learn and improve their skills. According to Barrot (2023), the tools appear to be capable of evaluating different aspects such as coherence, cohesion, language style, grammar, syntax, etc. The author also provides that ChatGPT's capacity does not seem to be limited to simple tasks, but learners may also utilize it for more difficult and complex assignments. Another aspect that ChatGPT has been proven helpful was outline generation. It has been shown that the tool can devise an outline into any formats required, such as a sentence, topic, or a paragraph, which would be essential for writing as structuring one's thoughts into a comprehensive and coherent essay is a challenge shared among students (Barrot, 2018; De Smet et al., 2012). However, scholars have also pointed out the cautions to be considered when students use the tool. Reliability seems to be the initial issue as its developers acknowledged the inaccuracy and unintelligibility in the tool's response. With regard to writing education, ChatGPT has also shown its limits in analyzing plagiarized texts and tailoring the writing to specific readership. Finally, relying solely on using the tool for writing, students may be subject to learning loss and difficulty in developing critical thinking skills, especially when they are not trained in advance. Arguably, much attention has been given to how to prevent students from relying on the tool while students' perception toward it has not been fully investigated.

It could be argued that while ChatGPT has become more and more popular among university students, research into how to incorporate the tool effectively to assist both students and practitioners is still necessary. Also, students' perception toward the tool is still rare in the literature. As such, this study sets out to fill the gaps, providing useful insights to both scholars and educators who are interested in incorporating ChatGPT into their teaching practice at the tertiary level in particular. The study proposes two research questions as follows:

1. With the instructor's training, to what extent is ChatGPT helpful in assisting students' English academic writing skills at the tertiary level in Japan?
2. With the instructor's training, what is the university students' perception of ChatGPT's effectiveness as a peer feedback tool after they use the tool for peer feedback?

Participants

There were 40 participants and they were first-year Japanese students who enrolled in an academic writing course of an English program at two private universities in Tokyo, Japan. The Cambridge proficiency test (Writing section) at the beginning of the semester indicated their proficiency levels ranged from A2 to B1 based on the Common European Framework of Reference for Languages (CEFR) framework.

Syllabus design

The 14-week syllabus required students to complete three writing assignments, one every three weeks. In the first week of each cycle, students learned key writing skills (e.g., topic sentences, unity, coherence), practiced in class, and received guidance on appropriate ChatGPT use (the free version). They then submitted a first draft using ChatGPT. In the second week, students participated in peer review and used ChatGPT again to revise their drafts. Afterward, the lecturer provided individual feedback for the final revision, which students completed without ChatGPT. The first assignment was a paragraph, while the second and third were short essays. Rather than emphasizing essay types (e.g., argumentative or compare-and-contrast), the syllabus focused on developing core academic writing skills. Since this study was part of a bigger project examining ChatGPT and teaching practices, only data during the first six weeks were utilized.

Data collection

The data collected for this study includes two main sets. The first set was the grades of the first and second draft to understand how useful ChatGPT was in assisting students’ academic writing. The second set of data was collected using questionnaires with both closed-ended questions and open-ended questions that let students share their opinions and evaluate the effectiveness of ChatGPT in their writing journey after the training. Due to the word limit, the full questionnaire was not included in this manuscript. After examining item-total correlations, the Likert scale demonstrated acceptable internal consistency (Cronbach’s $\alpha = 0.69$), approaching the threshold for reliability in exploratory educational research. Table 1 presents the descriptive analysis of the Likert Scale of the seven items.

Table 1. Descriptive analysis of the Likert Scale items

Item	1	2	3	4	5	6	7
M	3.95	4.19	3.95	4.33	3.90	3.19	3.33
SD	0.74	0.87	0.80	0.80	0.77	1.25	1.11

Since the primary goal of this study was to learn about how students perceived

ChatGPT and its role in helping students, grades from the final draft were not included, because the feedback used for it was provided by the lecturer, not ChatGPT.

Data analysis

The assignments were graded using adapted rubrics from Cambridge Assessment English, focusing on four main criteria: (1) content, (2) communicative achievement, (3) organization, and (4) language. The rubrics were modified to suit academic writing assessment. For instance, the organization criterion for a score of 15/25 was revised to “There are an introduction, body paragraphs, and a conclusion,” replacing the original, more general description. To avoid confusion, the grading scale was also changed from a 25-point to a 100-point maximum. Quantitative data were analyzed using SPSS 30.0, with a paired-samples t-test used to compare the first and second draft scores. Responses to open-ended questions were analyzed thematically to identify key recurring themes.

Results

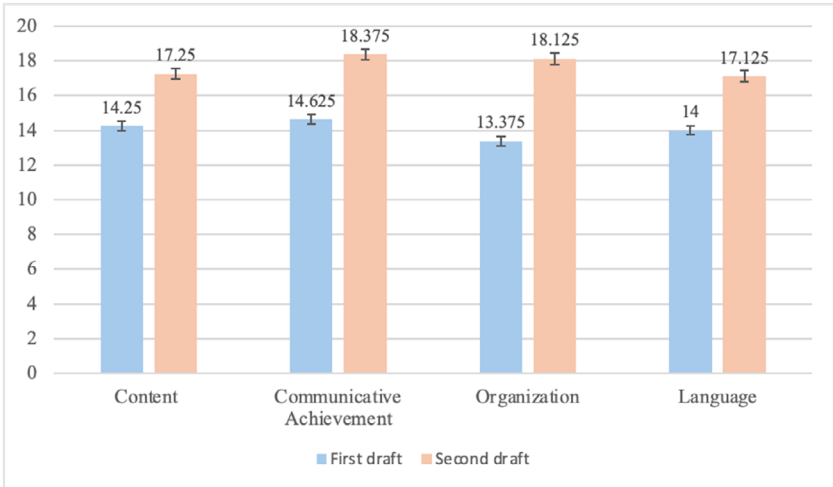


Figure 1. Comparison of mean scores between the first and second drafts

Participants’ second drafts showed clear improvement over their first drafts, with the average score rising from 56.25 to 70.7 out of 100. As illustrated in Figure 1, progress was evident across all four evaluation criteria, with the greatest improvement in organization ($M = 18.125$, $SD = 4.895$), followed by communicative achievement ($M = 18.375$, $SD = 3.078$), language ($M = 17.125$, $SD = 4.513$), and content ($M = 17.25$, $SD = 2.985$).

Table 2. Paired-sample t-test of writing scores before and after using ChatGPT

	95% confidence interval of the difference		t	df	Significance		Cohen's d
	Lower	Upper			One-sided p	Two-sided p	
Content	-4.074	-1.926	-5.649	39	<.001	<.001	3.359
CA	-4.991	-2.509	-6.111	39	<.001	<.001	3.881
Organization	-6.559	-2.941	-5.312	39	<.001	<.001	5.656
Language	-4.604	-1.646	-4.273	39	<.001	<.001	4.625

According to the paired-sample t-test (see Table 2), participants demonstrated significant improvement across all four criteria ($p < .001$ for all).

Participants’ impression of using ChatGPT to assist writing

Based on the Likert scale items and open-ended responses, most participants found ChatGPT’s feedback useful and effective for improving their writing and felt more confident about their second drafts. However, despite these improvements, they remained somewhat hesitant to use ChatGPT for future academic writing. While ChatGPT was perceived as effective, participants were uncertain whether it actually saved time during the writing process.

Table 3. Overall impressions of ChatGPT as a peer reviewer (open-ended question No. 2)

Theme	Number	Percentage (%)
Helpful, detailed, and encouraging feedback	28	56
Peer-like but limited human qualities	7	14
Overly positive and lacking criticism	3	6
Limited or inconsistent reliability	11	22
Effectiveness depends on user input	8	16

Table 2 presents the main themes regarding participants’ impressions of ChatGPT as a writing peer reviewer. Over half of the participants (56%) found ChatGPT’s feedback helpful, detailed, and encouraging. However, the remaining themes reflected more critical or negative views, with several participants noting its limited and sometimes inconsistent reliability:

Since ChatGPT usually provides correct information, it was actually hard to notice when it gave incorrect details, which made it difficult to judge the accuracy of the information.

Some participants noted that ChatGPT’s suggested revisions often lacked a human-like tone and tended to be overly positive when they sought more detailed or critical feedback. They also observed that the effectiveness of ChatGPT’s responses depended on how users formulated their input prompts:

It was difficult to create effective prompts, so sometimes the responses didn't quite match what I wanted to ask.

Table 4. Helpfulness of using ChatGPT for writing (open-ended question No. 1)

Theme	Number	Percentage (%)
Grammar and spelling correction	23	46
Improving sentence naturalness and style	4	8
Structural and content coherence	10	20
Providing suggestions and multiple options	9	18
Limitations and evolving use	4	8

Five themes emerged from the first open-ended question on participants' perceived benefits of using ChatGPT (see Table 3). Nearly half of the participants (46%) highlighted its usefulness for grammar and spelling correction, followed by structure and content (20%). The first three themes aligned with the evaluation criteria, whereas providing suggestions and multiple options represented a newly emerging theme. For example, one participant explained:

I received several suggestions and chose the one that best expressed what I wanted to say. It was very helpful.

Some participants also held a critical view of using ChatGPT, pointing out its limitations during the writing process:

At first, it wasn't very helpful. It just said "no problem" until I pointed out unclear parts. But as I used it more deeply, I realized it examined the meaning and grammar of each word in context, which was very useful for fine-tuning.

Table 5. Challenges or difficulties of using ChatGPT

Theme	Number	Percentage (%)
Difficulty getting desired or accurate answers	22	44
Prompting and input challenges	7	14
Loss of ownership or authorial voice	5	10
Limited or inconsistent reliability	7	14
No major difficulties / neutral	9	18

Table 4 summarizes the main themes related to the challenges participants encountered when using ChatGPT. About 18% reported no major difficulties or expressed a neutral attitude, while the remaining responses largely overlapped with the themes in Table 3. The most frequently mentioned challenge was the difficulty of obtaining desired or accurate answers.

Discussion and conclusion

Overall, although data were only collected during six weeks since this was a pilot study for a larger project, insights provided were helpful to both researchers and practitioners who are interested in using the tool in their practice. For starters, results indicated that with detailed and proper instructions, ChatGPT could be a helpful tool assisting students outside of class hours, fostering their autonomy. In particular, analysis of students' grades has shown that it could be utilized to help students structure their paragraph in a coherent manner, reducing the instructor's workload in giving individual feedback. At the same time, in the questionnaire, students also acknowledged its affordances in helping them revise their writing. In particular, ChatGPT was observed as a helpful tool in giving feedback on grammar, sentence structure, and content of the submitted writing. Secondly, in terms of students' perception of ChatGPT' peer feedback, questionnaire data indicated that while students acknowledged the feedback generated by the tool, they realized the tool's limitations. For instance, while they thought the tool could give detailed and encouraging feedback, creating a peer-like conversation for them to share, they also suggested that the feedback was overly positive and lacked criticism. Students also provided that the tool sometimes changed the tone, style, and suggestions of the original version, making it difficult to preserve the author's intent.

While the study succeeded in understanding ChatGPT's role in assisting students' writing and their perception of the tool, there are certain limitations to be acknowledged. First, since data were collected over a short period of time, students' competence in composing essays with the help of ChatGPT was not observed. Therefore, after this study, the authors will start to collect data in which students' writing is observed in a whole semester to corroborate the current findings. Also, this study requires an extensive amount of time from teachers at the outset as students need to be guided on how to use ChatGPT, provided detailed instructions after each writing draft, and afforded constant follow-ups during the feedback time. While this contributed to improving students' competence in using ChatGPT, it could also create more burden on the teacher's side. For this reason, university teachers should contemplate this syllabus carefully before incorporating it into their practice. Finally, there is a risk of students relying on the tool solely to finish their required assignments, creating more pressure on the teacher's side during the feedback and evaluation process. As such, it is recommended that teachers need to be confident that they understand their students' writing competence before the tool is introduced.

References

- Barrot, J. S. (2018). Using the sociocognitive-transformative approach in writing classrooms: Effects on L2 learners' writing performance. *Reading and Writing Quarterly*, 34(2), 187–201. <https://doi.org/10.1080/10573569.2017.1387631>
- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>

- Chaudhry, M. A., & Kazim, E. (2022). Artificial Intelligence in Education (AIED): A high-level academic and industry note 2021. *AI Ethics*, 2, 157–165. <https://doi.org/10.1007/s43681-021-00074-z>
- De Smet, M. J., Brand-Gruwel, S., Broekkamp, H., & Kirschner, P. A. (2012). Write between the lines: Electronic outlining and the organization of text ideas. *Computers in Human Behavior*, 28(6), 2107–2116. <https://doi.org/10.1016/j.chb.2012.06.015>
- Kexin, L., Yi, Q., Xiaou, S., & Yan, L. (2020). Future education trend learned from the Covid-19 pandemic: Take artificial intelligence online course as an example. In *2020 International Conference on Artificial Intelligence and Education (ICAIE)* (pp. 108–111). IEEE. <https://doi.org/10.1109/ICAIE50891.2020.00032>.
- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/003368822311628>

Author biodata

Bao Nguyen holds a PhD degree in International Cultures and Communication Studies and his major is TESOL. His research field is applied linguistics, especially in Teacher Psychology, CALL, and TESOL. His current interest is in language teachers' well-being and the importance of CALL in language teachers' professional development in the two countries.

Jiale Lu holds a PhD degree in Applied Linguistics from Waseda University. Her research interests focus on language learning motivation, learner identity, and technology-assisted language learning. She aims to integrate empirical research findings into pedagogical practices and address real-world challenges in language education.