Enhancing L2 Writing with ChatGPT for Japanese EFL Students

Emi Hennessy
University of Fukui

Reference Data:

Hennessy, E. (2025). Enhancing L2 writing with ChatGPT for Japanese EFL students. In B. Lacy, M. Swanson, & P. Lege (Eds.), *Moving JALT Into the Future: Opportunity, Diversity, and Excellence*. JALT. https://doi.org/10.37546/JALTPCP2024-04

This paper presents a preliminary study on the impact of generative AI, specifically ChatGPT, on the English essay writing of first-year engineering students at a regional national university in western Japan. The researcher investigates the potential of ChatGPT to improve task design for effective English learning. Addressing a gap in existing research, the research involved 27 students who revised their in-class English essays with the assistance of ChatGPT five times throughout the semester. The effects were assessed through pre- and post-tests of writing quality (linguistic aspect) and questionnaires (affective aspect). The results indicated positive outcomes in both linguistic and psychological dimensions. The findings offer significant insights into the effective and ethical use of generative AI in future classroom practices and provide reflections on potential challenges and directions for future research.

本稿では、西日本にある国立大学の工学部一年生が、英語のエッセイ・ライティングに生成AIを活用することによる言語面および情意面への影響に関する予備的研究について報告する。本研究では、ChatGPTの活用により効果的な英語学習の課題設計ができるかについて検証を行った。27名の学生が授業中に書いた英語のエッセイを、ChatGPTを活用して各々編集する課題を学期中に計５回課し、その取組の成果を言語面はAI活用の事前事後における英語エッセイの質の比較により、情意面はアンケートの実施により検証した。その結果、言語面および情意面ともに正の影響が確認できた。これらの結果により、今後の授業における生成AIの効果的かつ倫理的な活用について有益な示唆が得られ、今後の研究における課題や留意点に関する深い考察が提供される。

The rapid advancement of technology, particularly since the launch of ChatGPT in November 2022, has sparked global debate over its role in language education (Lo et al., 2024). This issue is also relevant in Japanese higher education, where integrating generative AI tools remains a challenge due to unclear national and institutional policies (MEXT, 2022; Praphan & Praphan, 2023). While some institutions, such as Sophia University, restrict the use of generative AI in coursework, others, like Ritsumeikan University, encourage its use in language learning (Sophia University, 2023; Ritsumeikan University, 2023). This divergence highlights the need to explore effective approaches for incorporating generative AI, such as ChatGPT, into Japanese EFL classrooms.

Motivated by this ongoing debate, the researcher conducted a small-scale study to investigate the pedagogical value of using ChatGPT in first-year writing tasks. The study aimed to assess its potential to support writing development and enhance learning motivation, particularly as students increasingly turn to AI tools in academic contexts.

Context

The participants in this study were first-year engineering students enrolled in general English courses at a regional national university in western Japan. Based on their learner profiles and observations of students in the same course taught in previous years, these learners have typically exhibited A1–A2 proficiency levels on the CEFR (Common European Framework of Reference for Languages) scale (Council of Europe, 2025) and have consistently encountered persistent challenges, including low confidence, limited motivation, and minimal classroom engagement. Many students tended to rely on machine translation tools, and in some cases, the use of generative AI was also observed. This reliance often occurred without fully engaging in the writing process, which may have impeded their language development.

In this paper, the term “generative AI” refers specifically to OpenAI’s ChatGPT, excluding other comparable platforms or applications. This study employed ChatGPT-3.5, the most recent free version available at the time, to examine whether structured use of the tool could support the development of students’ writing competence and motivation. The guiding research question was: To what extent does

the use of ChatGPT support writing competence and motivation among low-proficiency learners?

Overview of Previous Research

Although interest in generative AI in education is growing, empirical research remains limited and context-specific. Existing studies fall into three general categories: exploratory studies in Japan highlighting preliminary classroom applications of ChatGPT (e.g., Miyazoe, 2024), tool-focused research on AI integration (e.g., Yuasa & Takeuchi, 2024; Lambacher et al., 2022), and international case studies showing potential benefits in EFL writing instruction (e.g., Marzuki et al., 2023; Song & Song, 2023). While these studies offer valuable insights into the promise of AI-enhanced language learning, they often involve advanced learners, instructor-reported outcomes, or limited classroom implementation. For instance, Miyazoe (2024) conducted a preliminary study at a Japanese science university, highlighting the widespread use of generative AI tools such as ChatGPT for English language correction and proposing several strategies for incorporating generative AI into English writing tasks. Although the institutional context and course design in Miyazoe's study are comparable to the present research setting, the student populations differ significantly. Learners in the previous study were already active users of generative AI, whereas in the current setting, only 3 out of 27 students reported any prior experience with such tools for language learning. This disparity raises questions about the transferability of the findings to less AI-experienced learners. Furthermore, although the study presented the course design, it did not include evidence of classroom implementation or report measurable learning outcomes, leaving its practical impact unexamined.

In the global context, Marzuki et al. (2023) investigated the use of generative AI in Indonesia and reported potential improvements in student writing; however, these findings were based solely on instructors' perceptions and did not include direct assessments of student performance. By contrast, Song and Song (2023) conducted a mixed-methods study in China involving EFL learners who used ChatGPT as a supplementary tool. Their findings revealed improvements in writing quality and learner motivation; however, their participants possessed more advanced English proficiency and were evaluated using IELTS tasks. These are conditions that differ markedly from the present study's context, where first-year engineering students engage in non-academic, foundational writing tasks. Given the differences in learner profiles, task types, and assessment measures, the applicability of these findings to beginner-level EFL learners with limited AI experience remains uncertain.

As a result, there is a notable scarcity of studies examining the use of ChatGPT in basic-level writing courses for university students in Japan. Given the potential for students to rely on these "convenient tools" without fully engaging in the learning process, it is essential to critically investigate whether generative AI can deliver measurable educational benefits. Particularly in contexts where students have limited prior exposure to generative AI, further empirical research and practical experimentation with these technologies are necessary to better understand their potential impact on language learning outcomes in foundational EFL settings.

Methods

Class and Participants

This study was conducted in a compulsory general English course taught by the author from April to July 2024 at a regional national university in western Japan. The course, designed for engineering majors, met twice weekly for 90 minutes across 30 sessions and awarded two credits upon completion. Instructors had flexibility in course design, except for a final presentation requirement. The curriculum emphasized general English using all four skills, and textbook selection was left to each instructor.

A total of 27 students (26 male, 1 female) were enrolled, with English proficiency ranging from A1-A2 levels on the CEFR. According to a student profile questionnaire, three students had passed Eiken Grade 2, three had passed Pre-2, and three had passed Grade 3, levels generally corresponding to CEFR A1-B1 (Eiken Foundation of Japan, n.d.). The remaining students, whose proficiency levels did not exceed those of the Eiken passers, reported no experience with standardized English proficiency tests. Many students expressed low confidence in their English ability; however, 48.1% indicated that they found learning English enjoyable. When asked about the future use of machine translation or generative AI tools in English learning, 92.3% expressed interest, and 88.5% indicated a desire to learn how to use these tools effectively, highlighting the relevance of this study.

Course Design

The course primarily emphasized face-to-face instruction focused on speaking and listening, using an integrated-skills textbook (*Keynote 1A: Combo Split with My Keynote Online*; Bohlke & Cevik, 2016) designed for CEFR A1–A2 learners. In the third lesson, following the textbook's introduction, students were asked to express their own experiences and thoughts in English on the given theme by writing an essay using only

a dictionary during the final 30 minutes of class. The fourth class was conducted in an on-demand format. Students completed a ChatGPT-based assessment task, prompting the AI to review their essays (see Appendix A for the provided prompt). They then revised their work based on the feedback and rationale provided. Newly learned vocabulary, grammar, or expressions were highlighted in red by the students. Using the free online text-to-speech tool *Ondoku-san* (2025), students converted their revised essays into audio, which they used to practice and prepare for a verbal presentation in the next face-to-face session. The final version of the edited essay was submitted to the instructor. In the fifth class, students presented their prepared content in pairs. This five-step cycle was repeated five times throughout the semester, resulting in five writing assignments. In two of the five cycles, students were also required to submit a video recording of their practice presentation. The instructor reviewed all submissions, providing feedback on issues missed by the AI or content that exceeded students' proficiency levels.

Data Collection

Data sources included written assignments and questionnaire responses collected throughout the course for analysis. All writing assignments were collected either as in-class handwritten work or as digital submissions via Google Classroom. These included pre-test essays, five in-class writing assignments, five on-demand English writing revision tasks, and post-test essays.

For the pre- and post-tests, students wrote an English essay on a specified theme within 30 minutes using only a dictionary. To control for theme familiarity and difficulty, two essay topics were used: Theme 1: A person I respect; Theme 2: Why I chose engineering. Students were randomly assigned to one of the themes for the pre-test and wrote on the alternate theme in the post-test. This crossover design minimized potential bias (Tavakoli & Foster, 2011; Porte, 2010). All essays were anonymized but linkable for analysis.

Three questionnaires were administered during the course. The first, completed on paper at the beginning of the semester, was designed by the instructor to gather students' background information and previous use of translation tools or AI (multiple-choice) and their impressions of AI-assisted English learning (short-answer). The mid-term and final questionnaires were administered via Google Forms. Students completed them anonymously. These instruments explored students' experiences using AI to edit their English writing and their impressions of such technologies. All questions and responses were in Japanese.

Participants were fully informed about the study's purpose, procedures, use of findings, and data privacy. They provided informed consent, and the project was approved by the university's IRB (Institutional Review Board).

Results

The effects of using generative AI in English writing revision tasks were analyzed with an emphasis on both linguistic and affective aspects. The impact on the linguistic dimension was evaluated by comparing the pre-test of English writing conducted at the beginning of the semester with the post-test conducted toward the end of the semester. The impact on the affective dimension was analyzed based on the results of questionnaires conducted in the mid-term and at the end of the semester.

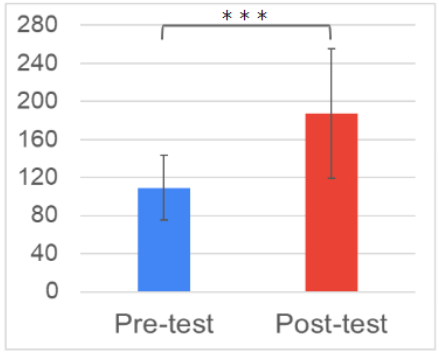
Linguistic Aspect

To evaluate the linguistic impact of the intervention, pre- and post-test results were compared based on word counts and rubric-based evaluations. Word count analysis was used to assess writing fluency, while writing proficiency was evaluated using the "6 Traits of Writing" rubric (Bridges, 2018). These measures were also used to explore the specific effects of using generative AI in English writing revision tasks. Of the 27 students, only 26 sets of writing samples were analyzed, as one student wrote on a topic different from the assigned post-test prompt, rendering the sample invalid.

The average word count within the 30-minute writing period increased from 109 in the pre-test to 187 in the post-test (Figure 1), indicating a notable gain in writing fluency. While this improvement may be partially attributed to intensive writing practice throughout the course, it is also plausible that the use of generative AI supported students' writing development. However, further research is needed to determine the specific extent of its impact.

Figure 1

Average Number of Words Written by Students for Pre- and Post-Test



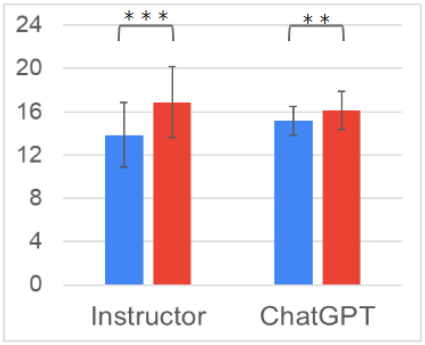
The 6 Traits of Writing rubric evaluates Ideas, Organization, Voice, Word Choice, Sentence Fluency, and Conventions on a 5-point scale. Two English instructors, one Japanese and one native English speaker, and ChatGPT-4 evaluated each essay. While students used ChatGPT-3.5 during the course, ChatGPT-4 was used for analysis, as it was the most recent version available at the time. Assessors scored each of the six traits (1 = Needs Improvement; 3 = Adequate; 5 = Excellent), with a total possible score of 30. To ensure objective assessment, all pre- and post-test essays were anonymized and randomized. Generative AI evaluated the essays twice, using the same prompt but with a different randomized order each time (see Appendix B for the prompt used).

Two reasons support the inclusion of generative AI evaluation. First, as seen in systems such as the ETS TOEFL iBT writing test, it is becoming common to combine automated scoring systems with human raters (ETS Japan, n.d.). Second, recent research suggests that "the qualitative feedback analysis indicated that both ChatGPT-3.5 and 4 consistently provided more relevant feedback on the EFL essays than the teacher raters" (Li et al., 2024).

The average scores from both the instructors and the generative AI were then analyzed (Figure 2). For the instructors' evaluations, the average pre-test score was 13.85/30, while the post-test score was 16.87/30. Participants' scores significantly improved from the pre-test (M = 13.8, SD = 2.9) to the post-test (M = 16.9, SD = 3.2), $t(25) = 4.15$, $p < .001$, d = 0.81. Similarly, the generative AI evaluation yielded a pre-test average of 15.15/30 and a post-test average of 16.12/30. Participants' scores also improved significantly from the pre-test (M = 15.2, SD = 1.3) to the post-test (M = 16.1, SD = 1.7), $t(25) = 2.93$, $p < .01$, d = 0.57, although the instructor-based evaluations showed a larger effect size (d = 0.81) compared to the AI-generated scores (d = 0.57).

Figure 2

Average Scores from Human Assessors and ChatGPT Based on "6 Traits of Writing" Rubric



In addition to evaluating overall writing proficiency, scores for each trait were analyzed individually (Table 1). Statistical analyses revealed significant improvements ($p < .05$) across all six traits from pre-test to post-test, with medium to large effect sizes observed. Notably, the greatest gains were found in Ideas (d = 0.79), Voice (d = 0.91), and Word Choice (d = 0.85), all indicating large effect sizes and reflecting substantial development in students' ability to generate content, express themselves clearly, and select appropriate vocabulary. Improvements in Sentence Fluency were also statistically significant, increasing from M = 2.02 (SD = 0.35) to M = 2.28 (SD = 0.39), $t(25) = 2.80$, $p = .011$, with a medium effect size (d = 0.55).

This suggests that students became more adept at constructing smoother and more varied sentences, thereby improving the readability and flow of their writing. Moderate gains were also observed in Organization (d = 0.42) and Conventions (d = 0.50), indicating statistically significant progress in structuring ideas and applying grammatical rules. However, these areas may benefit from more targeted or sustained instructional support.

Table 1

Descriptive and Inferential Statistics for Writing Traits Before and After Intervention (N = 26)

Trait	Pre *M* (SD)	Post *M* (SD)	*t*(25)	*p*-value	*d*
Ideas	2.91 (0.50)	3.41 (0.55)	4.04	<.001	0.79
Organization	2.64 (0.52)	2.91 (0.53)	2.15	.045	0.42
Voice	2.81 (0.42)	3.28 (0.47)	4.66	<.001	0.91
Word Choice	2.07 (0.29)	2.39 (0.32)	4.33	<.001	0.85
Sentence Fluency	2.02 (0.35)	2.28 (0.39)	2.80	.011	0.55
Conventions	2.05 (0.28)	2.21 (0.34)	2.57	.019	0.50

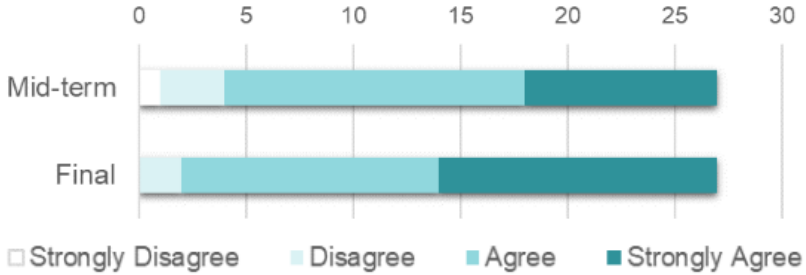
Affective Aspect

To understand how students perceived and felt about using generative AI in English writing revision tasks, I analyzed the results from mid-term and final questionnaires (see Appendices C and D), which were completed in Japanese by all 27 participants. The responses were translated into English by the author to report in this paper. The questions in both questionnaires were essentially the same; although individual responses varied slightly, overall trends were consistent.

In both surveys, 85.2% of students agreed or strongly agreed that using generative AI helped them acquire new vocabulary and expressions (Figure 3).

Figure 3

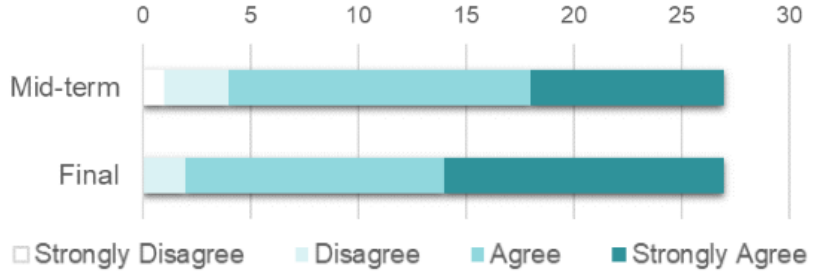
Survey Responses on Generative AI's Helpfulness in Learning New English Words and Expressions



When asked about grammar improvement, 85.2% of students responded positively at mid-term, rising to 92.6% in the final survey (Figure 4).

Figure 4

Survey Responses on Generative AI's Helpfulness in Improving Grammar Skills

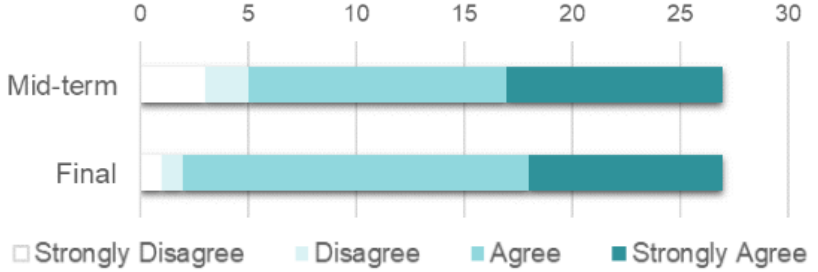


The proportion of "Strongly Agree" responses also increased over time, suggesting that students found AI more effective as they gained familiarity through repeated use.

Regarding confidence in their edited writing, 81.5% of students reported greater confidence at mid-term, increasing to 92.6% by the end of the semester (Figure 5).

Figure 5

Survey Responses on Generative AI's Helpfulness in Building Confidence in Edited Writing

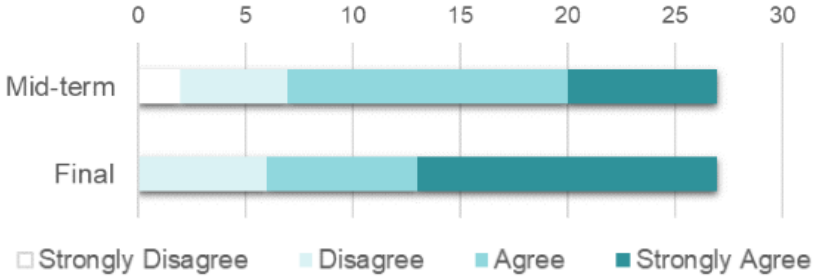


In terms of motivation, 74.1% of students agreed or strongly agreed at mid-term that receiving instant feedback from AI increased their motivation to learn. This figure rose

slightly to 77.8% in the final survey. Notably, no students selected "Strongly Disagree" in the final survey, and the number of "Strongly Agree" responses also increased (Figure 6).

Figure 6

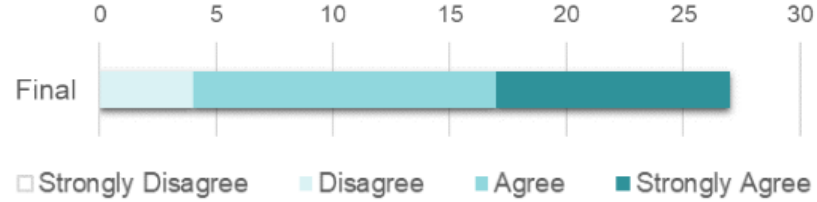
Survey Responses on Generative AI's Helpfulness in Increasing Motivation to Learn



One question included only in the final survey asked whether AI-assisted revision had helped improve students' overall writing skills. 85.2% (23 of 27 students) responded positively, with no "Strongly Disagree" responses recorded (Figure 7).

Figure 7

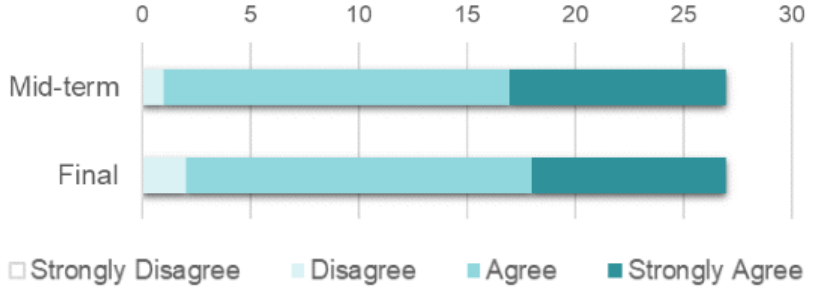
Survey Responses on Generative AI's Helpfulness in Improving English Writing Skills



Free comments revealed that students appreciated learning more natural expressions, appropriate vocabulary, and the rationale behind their errors. Several highlighted how the AI's feedback enhanced their understanding of English usage. When asked whether pair presentations of their edited writing helped consolidate learning, over 92% of students agreed or strongly agreed in both surveys (Figure 8).

Figure 8

Survey Responses on Pair Presentations Contributing to Consolidating Learned Knowledge



In summary, students viewed generative AI as a valuable tool for acquiring vocabulary and grammar, increasing writing confidence, and enhancing motivation. The writing-to-speaking cycle, particularly through peer presentations, further reinforced their learning, as reflected in their self-reported skill development.

Discussion

This study examined the linguistic and affective effects of using generative AI, specifically ChatGPT, in English writing revision tasks among EFL learners. The findings indicate that the integration of AI-supported feedback, combined with repeated writing practice, significantly contributed to improvements in both writing proficiency and learner engagement. These results are consistent with those of Tseng and Lin (2024), who found that incorporating ChatGPT into an EFL writing course enhanced students' writing performance and fostered "an engaging and interactive learning environment." Their study reinforces the potential of generative AI as an effective instructional supplement in language learning contexts.

In terms of linguistic outcomes, statistically significant improvements were observed across all six traits of writing, with medium to large effect sizes. A closer analysis of each trait suggests that different aspects of writing may have improved through different mechanisms. Traits such as Ideas, Voice, and Sentence Fluency likely improved due to the frequent and varied writing tasks assigned, which encouraged learners to develop fluency and expressiveness. Extended writing practice has long been recognized as a driver of fluency and idea development (Nation, 2009; Hyland, 2016), and these traits tend to

develop naturally as learners gain more experience expressing themselves in written form across diverse topics. Conversely, Word Choice, Conventions, and Organization appear to have been more directly influenced by the feedback provided by ChatGPT. Students not only received revised versions of their texts but also accompanying explanations for each revision, explanations that were explicitly designed to promote self-directed learning and metalinguistic awareness. This aligns with Li et al. (2024), who reported that ChatGPT-generated feedback was perceived as more relevant and instructive than traditional teacher feedback, particularly in explaining revisions. The statistically significant improvements in Conventions (d=0.50) and Organization (d=0.42), although moderate, support the conclusion that AI feedback played a meaningful role in helping students attend to grammatical accuracy and textual coherence, areas that typically require more explicit instruction and corrective feedback (Ferris, 2006; Bitchener & Ferris, 2012).

The affective data further underscores the value of generative AI in supporting learner engagement. High levels of agreement were observed in students' perceptions of ChatGPT's usefulness in helping them acquire new vocabulary and improve grammar. Notably, confidence in their writing and motivation to learn also increased by the end of the course. These findings are consistent with studies showing students value AI feedback for its immediacy, clarity, and nonjudgmental tone, which helps reduce anxiety and fosters motivation (Mohammed & Khalid, 2025; Zhan & Yan, 2025). Furthermore, the practice of having students present their revised writings in pairs likely contributed to knowledge consolidation, as over 92% of students reported that this activity reinforced what they had learned through revision. This reflects social constructivist views on language learning, where interaction and dialogue play critical roles in internalizing new knowledge (Vygotsky, 1978).

However, caution is warranted since these findings are based on a single cohort of 27 students from one course, limiting generalizability. It remains uncertain whether similar results would be replicated with different groups of learners, even within the same institutional context. Furthermore, the study did not include a comparison group that received writing instruction without generative AI support, making it difficult to isolate the specific contribution of AI integration from other factors such as task design or general writing practice. Therefore, while the observed improvements in writing competence and motivation are encouraging, the extent to which these outcomes can be attributed directly to ChatGPT use remains inconclusive. To address this limitation, a follow-up study is currently underway during the Fall 2024 term. This investigation involves two groups of first-year engineering students: one receiving targeted instruction on how to use ChatGPT-4 for English writing revision, and another following a comparable curriculum without specific instruction. This controlled comparison will allow for a more precise examination of the pedagogical impact of generative AI on both language development and learner motivation, offering further insight into its role within L2 writing instruction.

Despite the limitations noted, particularly the absence of a control group and the restricted sample size, this preliminary study provides encouraging evidence for the pedagogical potential of generative AI in supporting low-proficiency EFL learners. Overall, these preliminary findings provide a tentative but positive answer to the research question: To what extent does ChatGPT support writing competence and motivation in low-proficiency learners? The results indicate that ChatGPT contributed meaningfully and measurably to students' development in writing proficiency and to their affective engagement with the learning process. Consistent improvements in writing scores across all six traits, paired with students' overwhelmingly positive perceptions of AI-assisted revision, underscore the value of integrating generative AI tools into language learning contexts.

Beyond measurable gains in language proficiency, students also engaged critically with the AI tool itself. Survey responses revealed that through repeated use, many students developed an awareness of both the advantages and the limitations of generative AI. For instance, comments such as "I realized that not everything should be left to AI." and "I thought using AI would make me stop thinking and rely completely on it, which would lead to a decline in my ability to think. However, I now believe that AI can be very useful if used properly." reflect a deepening understanding of responsible technology use. Several students voiced concerns about over-reliance on AI and challenges in assessing the contextual appropriateness of its suggestions, reflecting issues noted in prior studies (Song & Song, 2023). These insights suggest that the intervention fostered not only linguistic growth but also critical reflection on information literacy and the ethical dimensions of AI use.

Conclusion

This preliminary study explored the potential of generative AI, specifically ChatGPT, as a tool to support English writing development among low-proficiency, first-year engineering students in a Japanese university setting. By integrating ChatGPT into repeated writing and revision cycles over a semester, the study found measurable improvements in students' writing fluency and proficiency across all six traits of writing.

These gains, confirmed by both human and AI evaluations, were particularly notable in traits such as Ideas, Voice, and Word Choice, suggesting that combining frequent writing practice with AI-generated feedback can meaningfully enhance learner output.

Likewise, the affective data indicated that students responded favorably to the integration of generative AI. Most reported increased confidence, vocabulary acquisition, and grammar awareness, along with heightened motivation to engage in English writing. The pair presentation component further supported knowledge consolidation and encouraged reflective engagement with their learning. Survey responses revealed that students not only improved their writing through AI assistance but also developed greater awareness of its limitations and ethical implications. This suggests an important ancillary benefit of the intervention: students' growing digital literacy and critical reflection on AI use.

Although limited by the absence of a control group and a small sample size, this preliminary study offers encouraging insights into the pedagogical role of generative AI in EFL writing instruction. The findings provide evidence that, when implemented with pedagogical intention, AI tools can function not simply as shortcuts but as scaffolds for language development and learner autonomy. Future research should aim to validate these findings in broader contexts and through more rigorous experimental designs. The ongoing follow-up study involving a comparison group will serve as a next step in clarifying the pedagogical role of generative AI and guiding its responsible integration into language education.

Bio Data

Emi Hennessy is currently an assistant professor in the Department of English Education, Center for Global Education and Research, at the University of Fukui. Her research interests center around effective and efficient English Language instruction for beginner-level learners, study abroad, and multilingualism and education in Japan. <ehenne@u-fukui.ac.jp>

References

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge. https://doi.org/10.4324/9780203832400

Bohlke, D., & Cevik, C. (2016). *Keynote 1A: Combo split with My Keynote Online*. National Geographic Learning.

Bridges, L. (2018). *Traits Writing: The Gold Standard of Writing Instruction and Assessment*. https://teacher.scholastic.com/products/ruth-culham-writing-program/pdf/traits_white_vFNL2.pdf

Council of Europe. (n.d.). Common European Framework of Reference for Languages: Learning, teaching, assessment – Table 1 (CEFR 3.3): Common reference levels: Global scale. *Council of Europe*. https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Eiken Foundation of Japan. (n.d.). 英検CSEスコアとは [*What is the EIKEN CSE Score?*] . https://www.eiken.or.jp/cse/

ETS Japan. (n.d.). 基本情報 [*Basic information*]. https://www.toefl-ibt.jp/educators/toefl_ibt/about/

Ferris, D. R. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81–104). Cambridge University Press. https://doi.org/10.1017/CBO9781139524742.007

Hyland, K. (2016). *Teaching and Researching Writing (3rd ed.)*. Routledge.

Lambacher, S., Kikuchi, H., & Obari, H. (2022). Exploring the impact of AI on EFL teaching in Japan. *Proceedings of the International CALL Research Conference, 2022*. 163–168.

Li, J., Huang, J., Wu, W., & Whipple, P. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities & Social Sciences Communications, 11*. https://doi.org/10.1057/s41599-024-03755-2

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. Y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments, 11,* 50. https://doi.org/10.1186/s40561-024-00342-5

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education, 10*(2). https://doi.org/10.1080/2331186X.2023.2236469

MEXT. (2022). 大学・高専における生成AIの教学面の取扱いについて [*Handling of generative AI in teaching aspects at universities and technical colleges*]. https://www.mext.go.jp/b_menu/houdou/2023/mext_01260.html

Miyazoe, T. (2024). Integrating ChatGPT into English education: Insights from a preliminary survey and course design at a Japanese science university. *Educational Studies, 66,* 49–65.

Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, *15*(7). https://doi.org/10.1186/s40468-025-00343-2

Nation, P. (2009). *Teaching ESL/EFL Reading and Writing*. Routledge.

Ondoku-san. (2025). *Text-to-speech web app.* https://ondoku3.com/

Porte, G. (2010). The impact of topic on ESL writing assessment: Evidence from a large-scale study. *Language Assessment Quarterly, 7*(4), 406–424. https://doi.org/10.1016/j.stueduc.2020.100911

Praphan, P. W., & Praphan, K. (2023). AI technologies in the ESL/EFL writing classroom: The villain or the champion? *Journal of Second Language Writing, 62*. https://doi.org/10.1016/j.jslw.2023.101072

Ritsumeikan University. (2023). 大学の英語授業に機械翻訳とChatGPTを組み合わせたサービスを試験導入 [*Pilot introduction of a service combining machine translation and ChatGPT in university English classes*]. https://www.ritsumei.ac.jp/news/detail/?id=3103

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14,* 1–14. https://doi.org/10.3389/fpsyg.2023.1260843

Sophia University. (2023). *Regarding ChatGPT and other AI chatbots (Generative AI)*. https://piloti.sophia.ac.jp/assets/uploads/2023/03/83431423966a7acb8457a729683dd1d8.pdf

Tavakoli, M., & Foster, P. (2011). The role of topic familiarity and task complexity in second language writing performance. *System, 39*(2), 176–186. https://doi.org/10.1515/applirev-2017-0017

Tseng, Y.-C., & Lin, Y.-H. (2024). Enhancing English as a Foreign Language (EFL) learners' writing with ChatGPT: A university-level course design. *Electronic Journal of e-Learning, 22*(2), 78–97. https://doi.org/10.34190/ejel.21.5.3329

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

Yuasa, M., & Takeuchi, O. (2024). Strategic use of machine translation: A case study of Japanese EFL university students. *AILA Review*, *37*(2), 215–240. https://doi.org/10.1075/aila.24020.yua

Zhan, Y., & Yan, Z. (2025). Students' engagement with ChatGPT feedback: implications for student feedback literacy in the context of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2025.2471821

Appendix A

A Prompt Provided to Students for Essay Revision Tasks Using ChatGPT-3.5

Instructions:

- As an American professional English teacher, your task is to provide corrections and suggestions based on the following constraints and input sentences.

Constraints:

- Ensure the text is easily understandable for someone with a TOEIC score of 500.
- Keep sentences concise.
- Correct any grammatical errors and highlight them in bold.
- Suggest more appropriate words or expressions using italic.
- Provide a list of reasons for each correction and suggestion.

Input:

- (Students pasted their typed essay here.)

Output:

Appendix B

A Prompt Used to Assess Students' Pre- and Post-writing Performance Using ChatGPT-4

Assess following writing based on the sample rubric:

Trait	Definition of each trait	5 (Excellent)	3 (Adequate)	1 (Needs Improvement)
Ideas	Develop a single topic to convey a clear message.	Clear, focused, and original. Well-supported ideas.	Some clarity and focus, but lacks detailed support.	Lacks clarity, focus, or supporting details.
Organization	Reveal ideas in a logical order, including a beginning, middle, and end.	Logical flow with strong transitions and structure.	Some organization but weak transitions or sequence.	Disorganized and hard to follow.
Voice	Increase reader engagement with intentional writer feeling, attitude, and emotion.	Strong, engaging, appropriate for the audience.	Inconsistent voice, lacks engagement.	No clear voice, dull or inappropriate tone.
Word Choice	Use the right word in the right place to convey just the right meaning.	Precise, descriptive, varied vocabulary.	Adequate but some overused or vague words.	Repetitive, vague, or incorrect word choice.
Sentence Fluency	Improve the readability of a piece with complete thoughts and sentence variety.	Smooth, varied sentence structures.	Some variety but some choppiness or awkwardness.	Very choppy, awkward, or repetitive sentence structure.
Conventions	Make writing more correct by following the rules of spelling, punctuation, capitalization, and grammar.	Few errors, does not interfere with meaning.	Some errors, but they don't overly distract.	Frequent errors that hinder understanding.

Topic the student was given:

- "A person or people I respect"
- "The reason I chose to study engineering"

The student's response:

- (The student's anonymized and randomized pre- and post-writing was pasted here.)

Appendix C

Questions in a Mid-term and an End-of-semester Questionnaire (Original)

The original survey was administered in Japanese. An English translation is provided in Appendix C for reference and transparency.

Note: The original survey included questions not only about the use of AI, but also about the course in general. However, only the questions related to AI usage—asked in both the mid-term and end-of-semester surveys—are presented here. Additionally, Question 6 was asked only in the end-of-semester version of the survey.

本授業における英語ライティング課題の編集時におけるAIツール使用に関するアンケート

以下の設問にご回答ください。本授業の英語ライティング課題の編集におけるAI等の使用について、現在までの状況や印象を聞かせてください。

1. 本授業の英語ライティング課題の編集においてAIを使用することが、新たな英単語や表現の習得に役立っている。
2. 本授業の英語ライティング課題の編集においてAIを使用することが、自身の文法力の向上に役立っている。
3. 本授業の英語ライティング課題の編集においてAIを使用することが、編集後のライティング課題に対する自信に繋がっている。
4. 本授業の英語ライティング課題の編集においてAIを使用することにより、瞬時に自分が書いた英文に対するフィードバックを受けられることが、学習意欲の向上に繋がっている。
5. 本授業の英語ライティング課題の編集においてAIを使用したあと、その内容をペアで発表することが、学んだ知識の定着に役立っている。
6. 本授業の英語ライティング課題の編集におけるAIの使用が、自身の英語ライティング力の向上に役立った。
7. 本授業の英語ライティング課題の編集においてAIを使用することについて、良いと思う点、学習に役立っていると思う点、難しいと思う点、改善点、質問、要望など、思うことを自由に書いてください。

Appendix D

Questions in a Mid-term and an End-of-semester Questionnaire (English Translation)

Survey on the Use of AI Tools for Revising English Writing Assignments in This Course

Please respond to the following questions. We would like to hear your current impressions and experiences regarding the use of AI tools in revising your English writing assignments.

1. Using AI to revise English writing assignments in this course has helped me learn new English words and expressions.
2. Using AI during the revision of writing assignments has contributed to improving my grammar skills.
3. Using AI in this course has helped build my confidence in my revised writing assignments.
4. Receiving immediate feedback from AI on my writing has increased my motivation to learn.
5. Presenting my revised writing in pairs after using AI has helped reinforce what I've learned.
6. The use of AI in this course has helped improve my overall English writing skills.
7. Please feel free to write your thoughts about using AI for revising writing assignments in this course, including what you found helpful or difficult, any benefits to your learning, suggestions for improvement, questions, or requests.