Reframing Hybrid Feedback in Japanese University EFL Writing: Complementary Roles of Peer Feedback and Generative AI

Hisami Tsuichibaru[1*]

[1]The University of Shiga Prefecture, Japan
[*]Corresponding author's email: kenchami2@gmail.com
[*]https://orcid.org/0000-0002-2196-7881
https://doi.org/10.54855/ictep.261110

® Copyright (c) 2026 Hisami Tsuichibaru

Received: 10/05/2026 Revision: 04/10/2026 Accepted: 07/10/2026 Online: 16/10/2026

ABSTRACT

The present classroom-based exploratory study re-examines hybrid feedback in a Japanese university EFL writing course for science majors. Thirty-six students used both peer feedback and ChatGPT-based feedback during a 12-week semester. Drawing on end-of-course questionnaires and a focused analysis of drafts, peer comments, AI feedback, and revisions, the study asked how students perceived the roles of peer and AI feedback and what pedagogical implications follow for hybrid feedback design. The findings suggested a complementary rather than competitive relationship. Peer feedback primarily supported content development, organization, reader awareness, and collaborative engagement, whereas generative AI primarily supported grammar, wording, and immediate local revision. When combined, the two sources appeared to encourage iterative revision and confidence; however, some students also experienced overload and difficulty prioritizing multiple comments. The paper argues that the value of hybrid feedback lies not in replacing human interaction with AI but in assigning distinct pedagogical roles to different feedback sources. It further suggests that the next stage of research should focus less on broad source comparison and more on comment quality, uptake, and AI-supported comment coaching. The study contributes a cautious classroom-based account of how generative AI can be integrated into EFL writing pedagogy in ways that preserve the educational value of peer interaction. Based on these patterns, the paper proposes a provisional four-stage framework for coordinating peer, AI, student, and teacher roles.

Keywords: generative AI, peer feedback, EFL writing, feedback literacy, revision

Introduction

Generative AI has altered the ecology of writing instruction with unusual speed. In higher education settings, students can now access instant linguistic suggestions, error correction,

CITATION | Tsuichibaru, H. (2026). Reframing Hybrid Feedback in Japanese University EFL Writing: Complementary Roles of Peer Feedback and Generative AI. *ICTE Conference Proceedings*, 11, 159-171. ISSN: 2834-0000, ISBN: 979-8-90148-780-8. DOI: https://doi.org/10.54855/ictep.261110

paraphrase options, and stylistic revisions with only a few prompts. This shift has sparked enthusiasm about efficiency and access, but it has also raised a more fundamental pedagogical question: if AI can supply immediate language-level feedback, what remains educationally distinctive about human interaction in the writing classroom? For EFL writing, where linguistic accuracy, rhetorical organization, and writer development are closely intertwined, the answer is unlikely to be a simple "AI versus teacher" or "AI versus peer" comparison. What matters is not which source wins in the abstract, but how different feedback sources function, what kinds of revision they invite, and what forms of learning they make possible.

The distinction among feedback roles is especially relevant in Japanese university EFL writing. Students in science-oriented programs often need to write summaries, short reports, project papers, and introductory research-style texts in English, yet many have had limited prior experience with extended writing or with giving and using feedback. In such contexts, teachers face a persistent tension between the need for individualized guidance and practical time limits. As a result, writing instruction can easily drift toward product correction, while opportunities for collaborative reading, evaluative judgment, and revision-oriented dialogue remain underdeveloped. This is precisely where peer feedback has long been valued: not simply as a low-cost substitute for teacher comments, but as a way of making writing social, dialogic, and reader-aware.

Despite growing research on automated and peer feedback, two gaps remain. First, broad comparisons and perception studies often leave the pedagogical division of labor among feedback sources underexplained (Tian & Zhou, 2020; Crosthwaite & Sun, 2025). Second, few classroom studies have examined peer and generative-AI feedback within the same revision cycle while tracing the types of change associated with each. Teachers therefore have limited evidence for deciding when to use each source and how students should reconcile competing suggestions.

To address these gaps, the present classroom-based exploratory study had three objectives: (1) to identify the roles students attributed to peer and ChatGPT-based feedback, (2) to examine the kinds of revisions associated with each source, and (3) to derive cautious design implications for complementary hybrid feedback. Section 3 presents the corresponding research questions.

The paper argues that hybrid feedback becomes pedagogically stronger when the functions and sequence of feedback sources are made explicit. Rather than treating peer and AI feedback as competing substitutes, the study interprets peer response as meaning- and reader-oriented support, ChatGPT as local language support, and teacher mediation as guidance for prioritization and judgment. The conclusion develops these findings into a provisional Complementary Hybrid Feedback Framework for future testing.

Literature Review

Peer feedback as collaborative and literacy-building practice

Peer feedback has been a major strand of second language writing research for several decades. Early studies emphasized the interactional and revision-related value of having students respond to one another's writing. Mendonça and Johnson (1994) showed that peer review sessions are not merely occasions for exchanging comments; they are sites of negotiation that shape later revision activity. De Guerrero and Villamil (2000) extended this line of work through a sociocultural lens, arguing that peer revision can activate mutual

scaffolding in the zone of proximal development. These studies are important because they frame peer feedback not simply as an instructional technique but as a learning process in which students develop through collaborative meaning-making.

Subsequent work has shown that the quality of peer feedback matters enormously. Tsui and Ng (2000) reported that peer comments can increase audience awareness and learner ownership, while Berg (1999) demonstrated that trained peer response affects both revision types and writing quality. Min (2005) further showed that training can improve the specificity and usefulness of student comments. Together, these studies suggest that the value of peer feedback depends not merely on participation but on whether students know how to give comments that identify problems, justify those judgments, and suggest workable revisions.

The importance of comment quality connects directly with the more recent notion of student feedback literacy. Carless and Boud (2018) define feedback literacy as the understandings, capacities, and dispositions students need in order to make sense of information and use it to improve work or learning strategies. From this perspective, peer feedback matters because it engages students in evaluative judgment. They do not just receive comments; they practice noticing quality, articulating criteria, and deciding how to act on feedback. In writing classrooms, this means that peer response can cultivate reader awareness, self-regulation, and metacognitive control over revision.

At the same time, peer feedback is not problem-free. Student comments may be vague, overly affective, excessively cautious, or linguistically limited. Differences in proficiency can shape both the production and uptake of peer comments, and students sometimes question the authority of feedback provided by classmates. Cheng and Warren (1997) showed that student attitudes toward peer assessment are not uniformly positive. These challenges do not invalidate peer feedback, but they do suggest that peer review requires training, structure, and mediation if it is to become a robust component of EFL writing pedagogy.

Automated and generative feedback in L2 writing

Automated writing feedback has a different history. Earlier automated writing evaluation systems promised speed, consistency, and scale, especially for grammar, mechanics, and surface-level editing. Warschauer and Grimes (2008) argued that such systems can reduce teacher workload and promote revision by providing immediate feedback. Their value lies not only in efficiency but also in the way they change the timing of revision: instead of waiting for teacher comments, students can address local issues immediately and repeatedly.

Recent generative AI tools have intensified both optimism and caution. Unlike earlier rule-based systems, large language models can provide context-sensitive suggestions, reformulate sentences, and generate apparently personalized advice. This flexibility makes them attractive in EFL writing classrooms, especially where students need support with grammar, phrasing, and sentence-level expression. At the same time, these tools raise questions about reliability, overdependence, and the pedagogical consequences of outsourcing judgment to AI.

Recent Japanese studies help clarify the issue. Allen and Mizumoto (2024) found that Japanese EFL learners often preferred ChatGPT to writing groups for editing and proofreading, especially when they wanted quick local assistance. Huang and Mizumoto (2024) reported that generative AI use was associated with gains in motivation and writing self-efficacy, while Huang and Teng (2025) showed that ChatGPT-generated feedback could support affective and behavioral engagement, even if it did not automatically guarantee deeper cognitive engagement. These studies matter because they show that students often

value AI, but not necessarily for the same reasons teachers or researchers do. AI is not simply a source of "better" feedback; it is a different kind of feedback.

A further point is that recent reviews suggest the field is moving rapidly but unevenly. Crosthwaite and Sun (2025) note that studies on generative AI and L2 written feedback have proliferated, but many still focus on broad outcome comparisons or on user perceptions. Comparatively fewer studies examine the internal mechanisms through which AI-related feedback is produced, interpreted, and acted on. This leaves an important gap for classroom research: rather than asking only whether AI improves writing, we need to ask what aspects of writing it supports, under what classroom conditions, and with what consequences for learner development.

Toward complementarity and AI-supported peer review

Research that directly compares or combines peer feedback and automated feedback provides a more nuanced basis for the present study. Chen and Cui (2022), for example, showed that peer feedback was more effective than automated writing evaluation for cohesion and coherence in continuation writing, suggesting that peers may be especially valuable for discourse-level and meaning-level issues. Tian and Zhou (2020) similarly demonstrated that learner engagement differs across automated, peer, and teacher feedback sources, and that uptake is shaped by individual and contextual factors rather than by feedback source alone.

More recent AI-supported peer-review studies shift the focus from AI as a direct evaluator of student writing to AI as support for student reviewers. Guo (2024) introduced EvaluMate, an AI-enhanced system designed to support the generation of qualitative peer comments. Guo, Pan, Li, and Lai (2024) showed that AI-supported peer review improved reviewers' feedback quality and writing ability. Guo, Zhang, Li, and Yu (2025) further examined how students revised their comments on peers' essays in an AI-supported environment, highlighting the potential of AI to scaffold feedback provision rather than merely correct texts.

AI-supported peer-review research is especially relevant to the current study because it suggests that the most pedagogically meaningful use of AI may not be to replace human response, but to improve the quality of human response. In other words, AI may be most valuable not when it becomes the primary source of feedback on writing, but when it helps students learn how to comment more concretely, more responsibly, and more usefully. That possibility aligns closely with feedback literacy research: if the goal is to develop students' capacity to make and use judgments, then AI should ideally strengthen rather than bypass that capacity.

The current study does not yet implement a full AI-supported comment-coaching intervention. However, it sits at the threshold of that design. By examining how students in one hybrid course experienced peer feedback and ChatGPT feedback, and by tracing which kinds of revisions appeared to be associated with each source, this paper provides a classroom-based foundation for moving from broad comparison to more mechanism-sensitive inquiry.

Recent practice-oriented studies in the requested publication venues provide a useful regional and applied context. Ha and Ho (2025) found that EFL postgraduate students valued both Grammarly and peer feedback, describing peer comments as more in-depth and constructive while viewing automated feedback as faster but sometimes inappropriate or limited. Truong and Tran (2024) reported broadly positive experiences with ChatGPT-supported writing practice, whereas Tran and Le (2025) showed why the reliability of AI-based grading should

be examined rather than assumed. These studies situate the present paper within a growing regional conversation about role, reliability, and classroom design.

Research Questions

1. How did students perceive the respective roles of peer feedback and generative AI feedback in a Japanese university EFL writing course?

2. What kinds of revisions appeared to be associated with peer feedback and with generative AI feedback, respectively?

3. What pedagogical implications follow for designing complementary hybrid feedback in Japanese university EFL writing?

Method

Context and participants

The study was conducted in a 12-week university-level English writing course for science majors at a public university in Shiga Prefecture, Japan.

The course aimed to help students develop academic writing through iterative drafting and revision. The participants were 36 undergraduate students aged 18 to 22. Their English proficiency, as indexed by TOEIC scores, ranged from approximately 450 to 700, representing a heterogeneous but generally intermediate-level EFL cohort. The class functioned as an intact group, and the study was conducted as a classroom-based inquiry rather than as a controlled experiment.

The analysis focuses on the course components in which students received both peer feedback and ChatGPT-based feedback on selected writing tasks. Other digital tools were used elsewhere in the course, but the present analysis is limited to these two components because they directly address the research questions. Participation in the questionnaire and permission to use classroom texts for research were voluntary. We removed student names and identifying details before analysis.

Course design and feedback procedure

The instructional design combined drafting, peer review, AI feedback, and revision. Students completed all focal drafting, peer-review, AI-feedback, and revision activities during scheduled class periods under instructor supervision. Students first produced an initial draft and exchanged it with classmates. Using a teacher-prepared rubric, reviewers commented on content development, organization, clarity, grammar, and vocabulary. The class used the same rubric and instructions. Students were not matched by TOEIC score, so the common procedure reduced—but did not eliminate—the possible influence of proficiency differences.

Students completed and submitted peer review before the class moved to the ChatGPT activity in each focal cycle. This in-class sequencing enabled the instructor to monitor the order of activities and kept ChatGPT outside the designated peer-review phase. Students then used ChatGPT-3.5, the version recorded for the April 2024 course protocol, with the same teacher-prepared prompt: “Please review the following English academic essay for grammar, vocabulary, and sentence structure. Identify errors, suggest corrections, and provide a brief explanation for each correction. Do not rewrite the entire essay; focus on specific issues and actionable feedback.”

Students received a brief task-specific orientation on using the common prompt, interpreting the output, and rejecting unsuitable suggestions; this was not extended prompt-engineering training. They compared the two feedback sources, revised their drafts, and submitted both the original and revised versions. The class discussed questionable AI suggestions to encourage critical evaluation rather than passive acceptance. Students' prior experience with and attitudes toward ChatGPT were not measured independently, so the perception findings may still reflect pre-existing attitudes.

Data sources

Two forms of data were used. The first was an end-of-course questionnaire with closed- and open-ended items about students' experiences with peer feedback, AI feedback, revision, motivation, and perceived usefulness. The second was a focused set of classroom artifacts: initial drafts, peer comments, ChatGPT feedback, and revised drafts from selected assignments. These artifact sets were used not to produce large-scale statistical claims but to identify qualitative patterns in the functions students attributed to each feedback source and the kinds of revisions that followed. The three questionnaire excerpts quoted in Section 5 are reproduced verbatim from the anonymized course records.

Because this paper presents an exploratory account of hybrid feedback in practice, it emphasizes triangulation rather than strong causal inference. Questionnaire responses provide insight into perceived roles and learner experience, while textual artifacts offer evidence of what students actually changed in response to available feedback.

Analytic approach

The closed questionnaire items were summarized descriptively, while open-ended responses were read repeatedly and coded for recurring themes. The coding focused on perceived usefulness, difficulty, confidence, collaboration, immediacy, and the distinctive affordances students attributed to peer feedback and ChatGPT. The analytic choice to foreground qualitative patterning rather than inferential statistical testing was deliberate. The study was conducted as regular teaching practice with an intact class, and the hybrid design was intended to support learning rather than to create fully separable experimental conditions. Students often moved recursively among drafting, peer review, AI use, and revision, so neat causal attribution was less appropriate than tracing recurrent pedagogical functions. For that reason, the present article treats the classroom as a site of patterned evidence rather than a laboratory experiment. We selected three open-ended responses to illustrate the peer, AI, and combined-feedback themes; they serve as examples rather than frequency estimates.

We analyzed textual artifacts qualitatively by comparing initial and revised drafts with the feedback students received. We examined revisions to determine whether they mainly addressed content development, organization, reader orientation, or local language issues such as grammar and wording. The study also reviewed peer comments and AI feedback for their functional emphasis. Rather than scoring feedback quality numerically, the analysis distinguished broad patterns such as content-focused suggestions, organizational suggestions, grammar-focused corrections, and reader-oriented comments. This exploratory analytic design was chosen because the paper seeks to clarify pedagogical roles and emerging mechanisms, not to claim definitive comparative superiority.

Findings

Peer feedback supported content development, organization, and reader awareness

A recurring pattern in the questionnaire responses was that students perceived peer feedback as especially helpful for content and structure. Many reported that classmates highlighted weaknesses in logic, paragraph flow, and the need for more concrete support. Representative comments included suggestions such as adding examples, clarifying an argument, or reorganizing paragraphs for stronger flow. These observations are pedagogically important because they indicate that students did not experience peer comments merely as error correction; they experienced them as responses from actual readers.

Several students also described peer review as useful not only because of the feedback they received, but because of the feedback they gave. Reading classmates' texts exposed them to other ways of organizing ideas and made them more aware of what readers notice. This finding is consistent with long-standing claims that peer response helps writers develop audience awareness and evaluative judgment. In the present classroom, this effect appeared in both the questionnaire comments and the revision artifacts: some revised drafts expanded explanations, added examples, or altered paragraph ordering in ways that aligned more closely with peer suggestions than with AI feedback. One student wrote, "Reading my peer's essay gave me ideas on how to make my introduction more engaging."

At the same time, peer feedback was not uniformly strong. Some students noted that classmates' comments were occasionally vague, overly brief, or insufficiently justified. A few also expressed uncertainty about the credibility of advice given by peers with similar or lower proficiency. These limitations did not negate the value of peer response, but they suggested that peer feedback worked best when students had at least some guidance on what to look for and how to phrase useful suggestions.

Generative AI supported language accuracy and rapid local revision

In contrast, students valued ChatGPT-based feedback mainly for language-level support. Students repeatedly highlighted its usefulness for identifying grammatical errors, awkward wording, and sentence-level problems that they had not noticed themselves. The immediacy of AI feedback was another major advantage. Several students appreciated being able to obtain quick comments without waiting for teacher response, and they often reported that AI feedback made revision feel more manageable.

Artifact analysis suggested that many local revisions—such as changes in article use, verb form, word choice, or phrasing—aligned closely with the kinds of issues ChatGPT flagged. In this sense, AI functioned as a local editor more than as a developmental reader. Students did not generally describe ChatGPT as helping them rethink their argument in deep ways. Instead, they described it as helping them "clean up" the language of their draft and making them more confident that their English was acceptable. A student explained, "The automated system showed me that I often misuse articles like 'a' and 'the,' which I hadn't noticed before."

However, this strength was matched by limitations. Students noted that AI feedback could be generic, excessively focused on minor language issues, or less helpful for argument quality, coherence, and persuasiveness. Some also worried that relying too heavily on automated feedback might weaken their own editing judgment. These comments indicate that students themselves recognized the limits of AI as a feedback source. Even when they valued it, they did not necessarily see it as sufficient.

Hybrid feedback promoted iterative revision, but also created overload

Perhaps the most important finding was that students often experienced peer feedback and ChatGPT feedback as complementary. In many questionnaire responses, students described the two sources as helping with different aspects of the same draft. A typical pattern was that peer comments pushed them to clarify ideas, provide examples, or improve organization, while ChatGPT helped them polish grammar and wording. One student summarized this division of labor clearly: "The automated system helped me fix grammar mistakes quickly, while my peer's comments gave me ideas on how to make my argument more convincing."

Together, the two feedback sources appeared to encourage a more iterative stance toward writing. Rather than treating revision as a single correction stage, students described moving back and forth among drafting, receiving comments, reconsidering ideas, and editing language. Some students who initially lacked confidence reported feeling more capable of revising after working with both kinds of feedback. In this sense, the hybrid design may have supported both performance and writer confidence.

Yet the combination was not automatically positive. Some students felt overwhelmed by the volume of feedback, especially when peer comments and AI comments pointed in different directions or varied in specificity. In those cases, the challenge was not a lack of feedback, but difficulty prioritizing it. This finding matters because it suggests that hybrid feedback is not effective simply because it is abundant. Students need help deciding what to act on first and why.

Summary of complementary roles

The findings are best understood as a differentiation of pedagogical roles rather than a competition among sources. Table 1 summarizes the recurring pattern that emerged across questionnaire comments and textual artifacts.

Table 1.

Perceived strengths and limitations of peer feedback and generative AI in the present study

Feedback source	**Main perceived strengths**	**Main perceived limitations**	**Pedagogical role**
Peer feedback	Content, structure, reader awareness, collaboration	Variable quality, vague comments, peer proficiency differences	Meaning-level and reader-oriented support
Generative AI	Grammar, wording, speed, local revision	Generic advice, weak on argument depth, risk of overreliance	Language-level support
Combined use	Iterative revision, complementary perspectives, confidence	Overload, difficulty prioritizing feedback	Sequenced hybrid support with teacher mediation

Discussion

From source comparison to complementary roles

The present findings suggest that hybrid feedback is most productive when feedback sources are treated as complementary rather than equivalent. Comparative questions remain useful, but they can obscure a pedagogically important point: different feedback sources do different work. In the present classroom, peer feedback was more closely tied to meaning,

organization, and reader awareness, whereas ChatGPT was more closely tied to local linguistic accuracy and rapid editing.

The meaning-level role attributed to peers is consistent with Chen and Cui (2022), who reported stronger peer-feedback effects on cohesion and coherence than automated writing evaluation, and with Tsui and Ng's (2000) account of audience awareness and learner ownership. The variability of peer comments also accords with Berg (1999) and Min (2005), whose studies demonstrate the importance of peer-review training.

The language-level role attributed to ChatGPT parallels Allen and Mizumoto's (2024) findings on editing and proofreading. Reports of confidence and engagement align with Huang and Mizumoto (2024) and Huang and Teng (2025), whereas overload and prioritization difficulties reinforce Tian and Zhou's (2020) conclusion that engagement and uptake depend on learner and contextual factors, as well as the feedback source. The present study extends these comparisons by locating peer and AI feedback within one classroom revision cycle.

Feedback literacy and the next step: from source to comment quality

A second implication concerns feedback literacy. If peer feedback matters because it develops evaluative judgment, then the next question is not only what source students use, but what kinds of comments they learn to produce. The present data suggest that students benefited from concrete, reader-oriented, revision-inviting comments. By contrast, vague or purely affective comments were less useful, and excessive quantities of undifferentiated feedback could create overload.

Recent AI-supported peer-review research provides a possible next step. Studies by Guo and colleagues (Guo, 2024; Guo et al., 2024, 2025) suggest that AI may be most valuable not as a direct assessor of finished texts, but as support for helping students generate better comments on peers' drafts. Read alongside the current findings, this research points toward a logical next step: future studies should examine comment quality, uptake, and the conditions under which AI can scaffold student reviewers' feedback literacy.

A mechanism-oriented view sharpens the contribution of a classroom report in a proceedings context. Practice papers often stop at whether students liked a tool or whether scores improved. Those findings matter, but they do not yet show why a design works pedagogically. By foregrounding the different functions of peer and AI feedback, the present study points toward a more cumulative line of inquiry in which classroom innovation is evaluated not only by uptake or satisfaction, but also by the kinds of judgment, reader awareness, and revision behavior it makes possible.

Implications for science-major EFL writing classrooms

The findings also carry practical implications for science-oriented EFL contexts. Students in such programs often need efficient support with language accuracy, but they also need practice explaining evidence, organizing claims, and anticipating readers. A feedback model that depends only on teacher correction is difficult to sustain, while a model that relies only on AI risks narrowing writing to local correctness. The present study suggests a more balanced alternative: reserve peer feedback primarily for meaning-level and reader-level concerns, use generative AI for language-level polishing and immediate support, and keep teacher mediation as the layer that helps students prioritize, interpret, and synthesize feedback.

The proposed balance neither romanticizes peer review nor overstates AI's precision. Rather, it calls for designing hybrid feedback around pedagogical roles. When roles are differentiated, hybrid feedback can support both writing quality and learner development. When roles are not explicit, the result may simply be more comments, not better revision.

Implications for feedback sequencing and teacher mediation

One additional implication concerns sequencing. The study did not experimentally compare alternative orders. However, the classroom procedure and students' descriptions support a sequencing hypothesis for future testing: peer comments may be used first for meaning and organization, followed by AI support for local linguistic polishing. Teachers can make this division of labor explicit and teach students to accept, adapt, reject, or defer suggestions. Teacher mediation is therefore essential when comments conflict or priorities are unclear.

Role differentiation is especially relevant for science-major EFL learners. Such students often value efficiency, clarity, and task completion, and many appreciate immediate language support because it reduces the burden of editing in a second language. Yet if AI support becomes the default center of revision, writing development risks narrowing to correctness. The present findings suggest a more productive balance: peer feedback can keep content and reader engagement visible, while AI can support the confidence needed to act on revisions. Hybrid pedagogy therefore works best when it protects the educational value of peer interaction rather than treating human feedback as merely slower machine output.

A provisional Complementary Hybrid Feedback Framework

Taken together, the findings support a provisional four-stage Complementary Hybrid Feedback Framework for classroom use: (1) in-class peer response foregrounds content, organization, and reader needs; (2) a constrained, common AI prompt supports subsequent grammar and wording revision; (3) students compare and prioritize suggestions rather than accept them automatically; and (4) teacher mediation addresses conflicts, questionable AI output, and higher-order concerns. The student remains the final decision-maker throughout the cycle.

The framework is a classroom design proposition derived from exploratory evidence, not a validated causal model. Future research should test its stages, alternative sequences, and effects on comment quality, uptake, writing development, and feedback literacy.

Limitations and Future Directions

Several limitations qualify the findings. First, the study involved 36 science-major students in one Japanese university course, so the results are exploratory and context-bound. Second, the analysis relied on questionnaire data and a focused set of classroom artifacts rather than a full corpus-based or controlled experimental design. Third, peer reviewers were not matched by TOEIC score, and proficiency differences may have influenced both feedback production and uptake. Fourth, the study did not measure students' prior experience with and attitudes toward ChatGPT independently.

Fifth, the AI findings are tied to ChatGPT-3.5 and may not generalize to newer models. Finally, peer and AI feedback were embedded in the same classroom revision cycle, and the artifact analysis was qualitative; revisions therefore cannot be attributed causally to a single source. The study also did not compare alternative feedback sequences or implement a systematic uptake-coding scheme.

These limitations point directly to future work. The next stage should examine the quality of student comments more systematically and trace which kinds of comments are actually taken up in revision. It should also test AI-supported comment coaching explicitly rather than inferring its potential from a hybrid classroom design. In addition, future research should explore the sequencing of feedback—when peer response should come before AI support, when the reverse may be useful, and how teacher mediation affects that sequence. Longer-term work should also investigate whether sustained hybrid feedback changes students' writing strategies, feedback literacy, and sense of writer agency over time.

Conclusion

The present classroom-based exploratory study re-examined hybrid feedback in Japanese university EFL writing and argued for a shift from effectiveness comparison to complementary role differentiation. In the present course, peer feedback appeared to support content, organization, reader awareness, and collaborative engagement, whereas ChatGPT-based feedback appeared to support grammar, wording, and immediate local revision. Together, the two sources often encouraged iterative revision and greater confidence, but they also risked overloading students when priorities were unclear.

The paper operationalizes complementarity through a provisional four-stage framework: in-class peer response for meaning and reader needs, constrained ChatGPT support for local language, student comparison and prioritization of suggestions, and teacher mediation. The framework keeps student agency and human judgment at the center of revision. Because it was derived from one intact class, treat it as a design proposition for future testing rather than a validated causal model.

Acknowledgement

The author thanks the participating students for allowing their classroom work and reflections to inform this study.

Compliance with Ethical Standards

The author reports no conflict of interest. Participation in the questionnaire and the use of classroom texts for research were voluntary. Student data were anonymized before analysis.

References

Allen, T. J., & Mizumoto, A. (2024). ChatGPT over my friends: Japanese English-as-a-foreign-language learners' preferences for editing and proofreading strategies. *RELC Journal, 57*(1), 89–106. https://doi.org/10.1177/00336882241262533

Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing, 8*(3), 215–241. https://doi.org/10.1016/S1060-3743(99)80115-5

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Chen, M., & Cui, Y. (2022). The effects of AWE and peer feedback on cohesion and coherence in continuation writing. *Journal of Second Language Writing, 57*, Article 100915. https://doi.org/10.1016/j.jslw.2022.100915

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education, 22*(2), 233–239. https://doi.org/10.1080/03075079712331381064

Crosthwaite, P., & Sun, S. (2025). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal, 57*(1), 207–219. https://doi.org/10.1177/00336882251386530

de Guerrero, M. C. M., & Villamil, O. S. (2000). Activating the ZPD: Mutual scaffolding in L2 peer revision. *The Modern Language Journal, 84*(1), 51–68. https://doi.org/10.1111/0026-7902.00052

Guo, K. (2024). EvaluMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing, 61*, Article 100864. https://doi.org/10.1016/j.asw.2024.100864

Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education, 63*, Article 100962. https://doi.org/10.1016/j.iheduc.2024.100962

Guo, K., Zhang, E. D., Li, D., & Yu, S. (2025). Using AI-supported peer review to enhance feedback literacy: An investigation of students' revision of feedback on peers' essays. *British Journal of Educational Technology, 56*(4), 1612–1639. https://doi.org/10.1111/bjet.13540

Ha, Y. N., & Ho, N. P. (2025). EFL postgraduate students' perceptions on the use of Grammarly and peer feedback to improve their academic writing skills. *International Journal of TESOL & Education, 5*(1), 25–49. https://doi.org/10.54855/ijte.25512

Huang, J., & Mizumoto, A. (2024). Examining the effect of generative AI on students' motivation and writing self-efficacy. *Digital Applied Linguistics, 1*, Article 102324. https://doi.org/10.29140/dal.v1.102324

Huang, J., & Teng, M. F. (2025). Peer feedback and ChatGPT-generated feedback on Japanese EFL students' engagement in a foreign language writing context. *Digital Applied Linguistics, 2*, Article 102469. https://doi.org/10.29140/dal.v2.102469

Mendonça, C. O., & Johnson, K. E. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly, 28*(4), 745–769. https://doi.org/10.2307/3587558

Min, H.-T. (2005). Training students to become successful peer reviewers. *System, 33*(2), 293–308. https://doi.org/10.1016/j.system.2004.11.003

Tian, L., & Zhou, Y. (2020). Learner engagement with automated feedback, peer feedback and teacher feedback in an online EFL writing context. *System, 91*, Article 102247. https://doi.org/10.1016/j.system.2020.102247

Tran, Y. V., & Le, T. H. G. (2025). Evaluating ChatGPT's reliability in grading writing assignments on the EOP learning platform. *ICTE Conference Proceedings, 7*, 1–19. https://doi.org/10.54855/979-8-9870112-8-7_1

Truong, C. Q., & Tran, T. M. H. (2024). The effectiveness of ChatGPT in students' practicing writing skills: A study at HCM UPES. *ICTE Conference Proceedings, 5*, 103–117. https://doi.org/10.54855/ictep.2459

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing, 9*(2), 147–170. https://doi.org/10.1016/S1060-3743(00)00022-9

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal, 3*(1), 22–36. https://doi.org/10.1080/15544800701771580

Biodata

Hisami Tsuichibaru is a part-time lecturer in Japan. Her research interests include AI-supported EFL writing, peer feedback, feedback literacy, and technology-enhanced language pedagogy.