Effects of interactive vs. complementary approaches to integrating automated and peer feedback on L2 writers' feedback quality and performance improvement

Da "Alex" Yan & Yu Gao

Published online: 06 Oct 2025.

Submit your article to this journal ↗

View related articles ↗

View Crossmark data ↗

Routledge
Taylor & Francis Group

Effects of interactive vs. complementary approaches to integrating automated and peer feedback on L2 writers' feedback quality and performance improvement

Da "Alex" Yan[a] ⬤ and Yu Gao[b]

aSchool of Foreign Language Studies, Wenzhou Medical University, Wenzhou, China; bSchool of Foreign Languages, Xinyang Agriculture and Forestry University, Xinyang, China

ABSTRACT

This study examined the impact of two approaches to integrating automated feedback (AF) and peer feedback (PF) on feedback quality and writing performance in a context of second language (L2) writing. The researcher designed two types of feedback interventions, where AF provided by Grammarly on draft writing products was either applied complementarily as an addition to PF on performance, or interactively embedded in the process of PF on performance and feedback messages. A total of 124 Chinese undergraduate EFL learners participated in the 16-week project consisting of three writing tasks, with 62 students in the experimental group (using the interactive approach) and 62 students in the control group (using the complementary approach). Mixed ANOVA tests were performed on PF quality and performance improvements measured across the three tasks. The results revealed that the interactive approach to feedback integration significantly enhanced students' abilities to effectively use assessment criteria and generate content-rich feedback comments. Additionally, the experimental group surpassed the control group in gaining performance improvement regarding organization. The research underscores the potential of the interactive approach to integrating AF and PF, and draws implications for researchers and educators interested in promoting the use of technology-enhanced PF in L2 writing.

1. Introduction

Recent advancements in research and practice regarding feedback for second language (L2) writing can be characterized by two key trends: (1) the proliferation of learner- and technology-based feedback sources; and (2) a shift towards formative and student-centered feedback approaches (K. Hyland & Hyland, 2019). A notable development is the increasing emphasis on peer feedback (PF) in L2 writing classrooms.

CONTACT Da "Alex" Yan ✉ alexyan1987@outlook.com 🏫 School of Foreign Language Studies, Wenzhou Medical University, Wenzhou, China; School of Foreign Languages, Xinyang Agriculture and Forestry University, Xinyang, China.

PF is widely recognized for its potential to promote writing development, evidenced by a meta-analysis showing a significant effect size of $d = 0.73$ (Vuogan & Li, 2023). It also helps reduce the instructional burden on teachers when scaling up feedback provision (Wu et al., 2022), and enhance students' agency (Cao et al., 2019), attitudes (Tian & Zhou, 2020), revision behaviors (Aben et al., 2022), and linguistic performance (Dao et al., 2023) in writing assessment and learning activities.

Conversely, PF can also be perceived as daunting due to its demands for higher levels cognitive and evaluative judgement skills (L. Cheng et al., 2023; García et al., 2024). Moreover, it could exert socio-affective impingements on learners such as low engagement and distrust among peers (Y. Zhang & Gao, 2024). A promising solution to enhance the efficacy of PF is to integrate it with diverse feedback sources (Elola & Oskoz, 2016; Loncar et al., 2023). Researchers have attempted the use of self-assessment (Cheong et al., 2023), teacher feedback/support (Ma, 2023; Wu et al., 2022), and PF from L1 learners (Niu et al., 2021) to complement L2 PF. Additionally, there is a growing trend to broaden the scope of feedback sources beyond human agents by incorporating automated feedback (AF) alongside PF, leveraging the logistical advantages of prompt responses from Automated Writing Evaluation (AWE) systems (Tian & Zhou, 2020; Z. Zhang & Hyland, 2022).

However, further investigation into the integration of AF and PF is still necessary. It is important to note that researchers have utilized various feedback design patterns when combining these two types of feedback (e.g. Xu & Zhang, 2023; Yao et al., 2021; Z. Zhang & Hyland, 2022). Unfortunately, previous empirical research has primarily focused on comparing the integration of AF and PF versus traditional modes of PF without AF support. To the best of the authors' knowledge, no studies have yet examined the effects of different integration approaches of AF and PF on L2 writing. This gap leaves unanswered questions about how to comprehensively understand the efficacy of using multiple feedback sources, particularly regarding how different feedback integration approaches influence L2 learning outcomes. More specifically, we are faced with the question of whether the combination of AF and PF, when interactively incorporated, or utilized in a complementary fashion, results in different performance improvements.

In light of the above, the current study aims to assess the effects of an interactive approach vs. a complementary approach to integrating AF and PF on the quality of PF comments and L2 writing performance improvements. This study contributes to the existing literature on multiple feedback sources in L2 writing pedagogy and has significant implications for researchers and educators seeking to advance contemporary feedback practices in L2 writing classrooms.

1.1. Supporting peer feedback: a quality-oriented perspective

PF refers to a formative type of peer assessment that provides qualitative information about peer performance, with the potential to promote learning improvement (Panadero et al., 2018). In recent years, significant progress has been made in understanding the complex and dynamic nature of PF. First, PF is a reciprocal and dialogic process (Myers & Buchanan, 2025). Therefore, successful translation from PF to actual learning gains calls for effective and meaningful interaction. For

example, Van Blankenstein et al. (2025) indicated that students would make better use of PF when then proactively request feedback during PF processes. Second, how students engage with the PF process and outcomes may vary depending on contextual and individual factors (Chen & Hu, 2025; Qian & Li, 2025). Thus, it is observed that PF is not always as effective as it is perceived (Noroozi et al., 2025), leading to a growing tendency in recent scholarly discourse to remain cautious about its potential shortcomings when designing and implementing PF. Among the various issues and challenges to PF's effectiveness, one central theme that stands out is its quality.

On the one hand, it is well established that the quality of PF is a critical precondition for learning improvement and active feedback uptake (Er et al., 2021; Zong et al., 2021). Students can reap multidimensional gains from quality PF. For example, high-quality PF provides not only a detailed understanding of students' current performance and learning issues but also suggestions and action plans for future improvements (Haro et al., 2023). Additionally, students receiving PF with higher quality are more likely to develop positive attitudes and motivations towards the active utilization of feedback (Patchan et al., 2016). On the other hand, the wider application and acceptance of PF are hindered by its alleged quality deficiencies. PF is often questioned for being superficial and lacking validity compared to teacher feedback, due to peers' limited domain knowledge and feedback literacy (Latifi et al., 2023). In L2 writing, a long standing criticism facing PF is its overemphasis on local issues rather than global issues in writing, thus diminishing its practical use as an instructional intervention (X. Cheng et al., 2023). Doubts about PF quality also represent a socio-affective issue, stemming from students' distrust in their peers' capabilities or an inharmonious learning atmosphere (Ibarra-Sáiz et al., 2020; Yu & Lee, 2016).

To promote the quality of PF, scholars and educators have invested considerable efforts in providing instructional scaffolding and material/technological support (Bürgermeister et al., 2021). First, instructional and pedagogical accommodations were implemented to prepare students with the necessary pre-knowledge for and procedural familiarities with effective PF practice. Therefore, researchers have provided PF training (Man et al., 2022), offered teacher interventions to correct or indicate erroneous/inaccurate PF comments (Sun & Wang, 2025), and enact dialogic PF to promote knowledge co-construction (Van Blankenstein et al., 2025; Zhao et al., 2024). Second, teachers provided material-based support as a cognitive facilitator to support students' agentic and constructive engagement with PF processes. For instance, researchers introduced regulation scripts to promote critical thinking in PF processes (Shao et al., 2023), and included students in co-creating the feedback rubric to enhance their ownership of the assessment task (Yan, 2024). Third, technology-enhanced platforms have been applied as communicative facilitators, enhancing dialogic PF and its uptake (Wood, 2022). Taken together, existing studies supporting PF quality either apply cognitive scaffolds that enable students to co-construct meaningful evaluative judgments or social scaffolds promoting deeper engagement with the PF processes.

In the present study, we designed and implemented a supportive approach that addresses both the cognitive and social aspects of PF by integrating AF and PF

sources into a dialogic PF intervention to enhance PF quality and its effects on writing performance. The rationale is twofold, aligned with the reciprocal nature of PF (Hovardas et al., 2014), where students involved in PF possess a dual-faceted identity as both feedback providers and recipients. First, the availability of alternative feedback sources increases the likelihood for students to receive a more comprehensive set of evaluative judgments on their writing. Second, the opportunity for peer learners to discuss or utilize AF enhances their chances of providing more constructive and in-depth PF comments.

1.2. Theoretical and empirical foundations for integrating feedback sources

With sustained changes in pedagogy and feedback practices (K. Hyland & Hyland, 2019), the introduction of alternative feedback sources into L2 writing classrooms provide students with more comprehensive feedback, offering varying degrees of explicitness and diverse feedback strategies (Niu et al., 2021). The integration of multiple feedback sources is supported by theoretical insights on self-feedback and the empirical evidence of the complementary use of feedback sources.

Self-feedback, or feedback generated by students themselves from external resources, plays a vital role in enhancing performance and learning (Lipnevich & Smith, 2022). According to Nicol (2021), self-feedback occurs when students actively engage with and effectively compare internal and external *frames of referenc*e, which broadly encompass various feedback sources, such as peer assessments, teacher comments, and self-assessments. By doing so, students develop a deeper understanding of their strengths and weaknesses, leading to improved self-regulation and academic performance. Lipnevich and Smith (2022) incorporated the above understanding of self-feedback into the Student-Feedback interaction model, asserting that: (1) feedback from various external sources is expected to elicit differential reactions from students; and (2) all external feedback must be internalized into self-feedback to actually promote changes in performance and learning. To support the generation of self-feedback, Nicol (2021) thus advocates for the creation of a learning environment where 'students' interactions with others and with resources' can increase the possibilities 'to turn natural feedback comparisons into explicit comparisons' (p. 768). In learning scenarios involving various feedback providers, an effective approach to support self-feedback generation would be to integrate feedback messages from different sources (Van der Kleij, 2024).

At the same time, L2 writing researchers have found that the characteristics of feedback from diverse sources are often complementary, while their combination can address the shortcomings of one type of feedback by leveraging the strengths of the other. For instance, Niu et al. (2021) found that teacher feedback and PF (from both L2 and L1 learners) complement one another well in terms of feedback points (85% were non-overlapping), focus (L1 peer and teacher feedback emphasized global issues, while L2 PF focused more on local issues), and strategies (peers tended to use suggestions, whereas teachers relied more on questioning). In another interventional study by Sun and Wang (2025), the researchers found that teacher's direct correction of erroneous PF enhanced adoption and revision accuracy rates, compared to students exposed to indirect indication of PF inaccuracies or direct usage of PF for revision.

Additionally, researchers have extended the scope of complementary sources of feedback to include AF. For example, Tian and Zhou (2020), through naturalistic observation in a multiple case study, highlighted that a multi-source feedback design in L2 writing enhanced learner engagement and writing development, as AWE was more effective at addressing surface-level issues (i.e. lexical-level), while PF tackled both surface and meaning levels (i.e. grammar and mechanics). Similarly, Tan et al. (2023) found in their factorial study that combining AWE and asynchronous PF was highly regarded, with students outperforming their peers in terms of complexity, fluency, and accuracy measures.

In summary, the theory of self-feedback underlines the necessity of incorporating feedback from diversified sources to enhance learning, while the complementary strengths of different feedback types support their integrative use in L2 writing pedagogy.

1.3. Integrating automated and peer feedback: two approaches

AF is a fast-evolving subdomains within assessment and feedback research. Over the years, our research interest has shifted from written corrective feedback automated by AWE systems (W. Liu, 2024) to AF provided by latest intelligent technologies, such as large language models and generative artificial intelligence (Chan et al., 2025; Yang & Chen, 2025). Recent research synthesis have highlighted the value of AF as a more logistically efficient source of feedback with sufficient validity and trustworthiness (Fan & Ma, 2022; Zhai & Ma, 2023). Moreover, similar to PF, studies show that student engagement with AF is subject to diverse array of factors, thus, integrating AF with human-led feedback, such as PF and teacher feedback, has been recommended to support sustained writing development and active feedback engagement (Li & Ke, 2025).

However, effective combination of AF and PF requires well-designed feedback integration patterns beyond simply providing two feedback sources simultaneously. Drawing on Nicol's (2021) concept of explicit comparison, effective integration should occur through comparisons and interaction between multiple feedback sources. As such, attention must be given to the internal feedback designs employed by researchers in relevant studies. Upon synthesis, two major approaches for integrating AF and PF are presented below.

The first, and more common, approach is the complementary combination of different feedback sources. Zhang and Hyland (2022) proposed and examined an integrated model that systematically utilizes three types of feedback: AWE, PF, and teacher feedback in L2 writing classes. In the 16-week writing course, students were required to complete two 800-word essays, following a 4-week process during which they received feedback from Pigai (a commonly utilized AWE system in China), peer learners, and the teacher, respectively, in each subsequent week after the initial drafting. The study revealed that this integrated model resulted in higher levels of cognitive, affective, and behavioral engagement, as well as deeper-level revision among its users. Several similar studies adopting experimental designs support these findings. For example, Liu et al. (2023) reported that students exposed to a peer-automated-combined-feedback condition outperformed those using a conventional AWE feedback approach in terms of writing performance, learning motivation, critical

thinking, and ability to control L2 writing anxiety. In Liu et al. (2023), learners in the experimental group first reciprocally commented on peer learners' writing artifacts using instructor-devised rubrics, and then revised their own writing based on feedback generated by both AWE and peers. A special case of the complementary application of PF and AF can be found in Xu and Zhang (2023), where AWE was applied three times throughout the writing and revision process: first as the primary source of feedback during the first round of revision, then as a complementary source after students received feedback from peers and the teacher in the second and third rounds.

The other approach to integrating AF and PF is to create opportunities for interaction between the two types of feedback. Yao et al. (2021) designed a feedback intervention that required learners to provide dialogic feedback on both peers' writing products and the Pigai-generated comments they received. The findings indicated that the inclusion of AF in dialogic peer assessment promoted a positive mentality and increased learning motivation among L2 writers. In a more recent study by Guo et al. (2024), researchers devised a different feedback intervention: feedback providers first generated feedback on writing, then sought evaluation of their feedback comments from an AI-based tool, and finally revised their original comments based on the AI's evaluation before offering the finalized version to the feedback recipients. Findings from Guo et al. (2024) revealed that this integrated feedback intervention significantly improved both PF quality and writing development. To better understand the interaction between PF and AF in the two above studies, we can refer to the three-step framework of dialogic PF proposed by Zhao et al. (2024). Specifically, in Yao et al. (2021), the students offered not only 'peer-feedback-on-performance' but also 'peer-feedback-on-automated-feedback'; while in Guo et al. (2024), the student and AI collaboratively engaged in a more sophisticated feedback process featuring the transition from 'initial-peer-feedback' to 'AI-feedback-on-peer-feedback' and then to 'Re-feedback based on the feedback-on-feedback'.

1.4. The current study

Based on the reviewed literature above, the following logic could be drawn: First, integrating PF and AF in L2 writing could promote the quality of PF messages in L2 writing. Second, an increase in PF quality could be translated into writing performance improvement. Finally, an interactive approach to integrating PF and AF is expected to have different effects on PF quality and writing performance improvement compared to the complementary approach of integration. Thus, the current study compared the effects of these two approaches (interactive vs. complementary) to integrating AF and PF in L2 writing classes on PF quality and writing performance development. The following research questions guided this study:

1. To what extent can the interactive approach to integrating AF and PF affect the quality of peer feedback on L2 writing, compared to the complementary approach?
2. To what extent can the interactive approach to integrating AF and PF affect L2 writing performance improvement, compared to the complementary approach?

2. Methods and materials

2.1. Participants

This study was conducted at a university in China during the spring semester of 2024. Participants were selected from four different classes of a 16-week 'Intermediate English Writing I' course, all taught by the same teacher and following the same curriculum and teaching plan. These classes implemented reciprocal and dyadic PF as a form of classroom-based formative assessment. Additionally, we sought to preserve the students' original dyadic partnerships by including only dyads whose members both agreed to participate in the project.

As a result, we recruited a total of 124 sophomore L2 students (M_{age}=19.8 years; 15 males and 109 females) as participants (30 from class 1, 28 from class 2, 32 from class 3, and 34 from class 4). The PF dyads were randomly assigned to two groups: (1) the complementary integration control group (CG; n=62), and (2) the interactive integration experimental group (EG, n=62). Based on a pre-project writing ability evaluation following the requirements of Test for English Majors band 4 (TEM-4), a national English proficiency test for English majors in China, there were no statistically significant differences in the TEM-4 writing grades among the two groups [$F(1, 122)$=1.23; p=.269; η^2_p=0.01]. The participants exhibited intermediate to intermediate-high levels of English writing ability, equivalent to B1 and B2 levels on the Common European Framework of Reference for Languages (CEFR). Additionally, we administered a survey on students' prior experiences with formative feedback, which showed that the participants frequently engaged in PF activities (M=3.82 on a five-point scale, SD = 1.14). However, only 7 students had prior experience with the combined use of AF and PF (as indicated by their answers to open-ended questions in the survey).

Research ethics, educational equity, and ecological validity (Kasprowicz & Marsden, 2018) were considered. First, all participants voluntarily took part in the research project and were fully informed of the research purpose, details, and privacy protection policies. Second, CG members and non-participants received all the training, materials, and instructions on the integrative use of AF and PF after the research project. Non-participants were exposed to the same instructional, practice, and assessment activities; however, the formative assessment activities for non-participants employed only dyadic PF, whereas participants used a combination of AF and PF. Third, all the writing artifacts produced during the project were used as part of the students' final grade for the course. Bonus marks were awarded for exceptional efforts in providing and making use of PF for writing development.

2.2. Intervention

The interventional design in this study involved two experimental conditions with different feedback models during the writing process (Figure 1). To maintain the fidelity of the intervention, we followed the 'feedback source → feedback messages → feedback processing' flow of student-feedback interactions outlined by Lipnevich and Smith (2022).
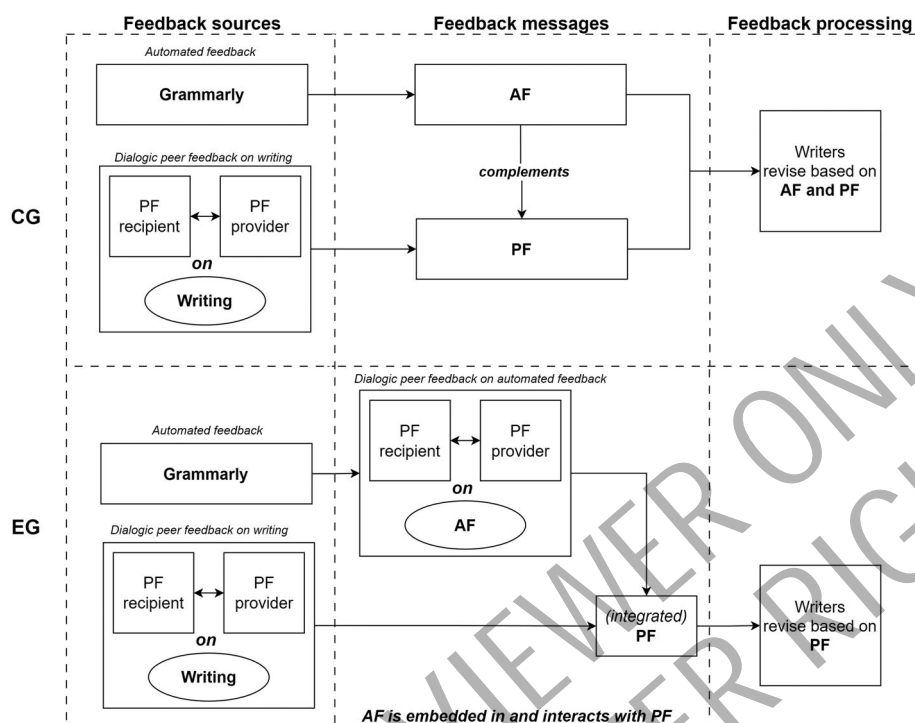
Figure 1. Feedback models in the two experimental conditions.

During all the feedback sessions, dialogic written PF was practiced. Following the three-step model proposed by Zhao et al. (2024), feedback recipients were required to give comments on the feedback they have received to the feedback provider; while the feedback providers, after discussion with the feedback recipients, were required to explain or reproduce their initial feedback comments to support the recipients' better understanding and utilization of the comments. Students have already had experience with such a mode of PF in a writing course in the previous semester. All PF comments as well as subsequent feedback-on-feedback and re-feedback messages were put in the discussion threads on Shimo.im (a Chinese online office application that supports collaborative editing of documents). To support effective and constructive PF activities, the first author hosted a three-session pre-semester online training and offered material scaffolds including prepared rubrics and regulatory scripts for the PF processes.

Both groups utilized the same feedback sources: AF provided by Grammarly and PF collaboratively generated by peer dyads through dialogic written discussions on writing performance. However, the integration of these feedback sources differed between the two groups. In CG, AF was treated as a complementary resource to PF. AF messages were excluded from the dialogic discussions during the PF process and remained accessible only to the L2 writers. Conversely, in EG, the PF process was extended to include discussions about both the original PF and the AF messages. This allowed participants to co-construct integrated PF messages during the dialogic (feedback-on-feedback) phase. Following these discussions, the original PF providers

refined their comments based on the consensus reached. After the feedback processes, L2 writers in the CG group received a refined version of PF along with the original AF comments for their subsequent revisions. In contrast, writers in the EG group were provided with an integrated PF refined by the original PF provider with reference to and consideration of AF messages. Additionally, EG writers could also refer to the original AF comments during the revision process, in case of need. All students in both groups completed their revisions individually.

2.3. Measures

2.3.1. Feedback quality scoring rubric

To assess PF quality, we adapted the feedback quality scoring rubric developed by Gielen and De Wever (2015). To better align with the dialogic PF approach used in the current study, we also referred to alternative feedback quality evaluation instruments, such as the coding scheme in Kerman et al. (2024), and readjusted the overall weight distribution as well as descriptors in several sub-categories (Appendix A). The adapted instrument encompassed three major categories: use of criteria; feedback contents; and feedback styles. Points across the three categories were summed up to represent the overall quality evaluation score of a PF message, with a maximum score of 100.

To ensure inter-rater reliability (IRR), the first author and four co-researchers independently coded 15% of the total PF messages collected from the study. IRR was measured using Fleiss' kappa: use of criteria (0.93); feedback contents (0.84); and feedback styles (0.88). Discrepancies in coding were solved when consensus was reached through discussion. Then, the coders split the remaining coding duties of feedback quality.

2.3.2. Writing tasks and grading

Focusing on the effects of feedback integration, this study measured L2 writing *performance improvement* in the tasks during the intervention (i.e. the progress students made after revising based on feedback, calculated as the revised version score minus the draft version score) rather than the overall scores.

Writing prompts were developed in accordance with the official guidelines for the TEM-4 test. The writing task required students to summarize a given passage and comment on specified points in no fewer than 200 words. The TEM-4 writing tasks were designed to assess four core writing competencies: (1) content relevance, (2) content sufficiency, (3) organization, and (4) language use. Prior to the project, the first author and co-researchers created or adapted 25 writing prompts to establish a task bank.

To select prompts with similar difficulties and topic familiarities, the first author recruited five experts in the field of L2 writing and educational assessment as Delphi panelists. First, panelists collectively clustered all the writing prompts in the task bank according to the topics and corresponding perceived level of topic familiarities; Then, 15 prompts with similar themes in societal issues were selected; Finally, each panelist rated the task difficulties on a 10-point scale and recommended three

prompts with similar task difficulties. After the Delphi process, six prompts were recommended to the first author for further consideration. The first author selected three prompts out of the six candidates through the readability assessment using ARTE (https://linguisticanalysistools.org/arte.html). The results showed that the chosen prompts had a mean Flesch readability score of 43.6, on par with official TEM-4 writing tasks and suitable for second year undergraduate L2 writers.

Nevertheless, the official TEM-4 scoring rubric was not used in this study to evaluate writing performance, as its impression-based design could potentially lead to subjective grading and low discrimination of student performance (Öztürk et al., 2019). Moreover, the official rubric employs a holistic grading approach, which precludes the reporting of writing performance across individual scoring dimensions. Therefore, we adapted the analytical rubrics created by Golparvar and Abolhasani (2022) and Zhang and Lu (2022). This rubric evaluated writing across three key dimensions: (1) content, (2) organization, and (3) language use, each assessed on a 5-point scale. The adapted rubric was suitable for the current study, as its assessment focus aligned with both the purpose of the research and the underlying writing performance dimensions within the official TEM-4 writing task scoring rubric.

The scoring process was conducted blindly, ensuring that raters were unaware of both the authors and the version (i.e. whether it was a draft or a revision) of the writing products. A similar approach to assessing IRR was undertaken to ensure scoring reliability. After the five coders individually grading 50 writing products, the following reliability coefficients were attained: content (0.82), organization (0.85), and language use (0.91). As sufficient grading reliability was established, the coders divided the remaining writing performance evaluations.

2.4. Procedures

This study employed a short-term longitudinal experimental design to assess the effects of different approaches to feedback source integration on learning over time. The research was embedded into the writing course, which was assigned 3 sessions per week (each 45 min). The course took place in computerized language laboratories where access to the Internet, digitalized language assets (e.g. parallel corpus, and compilations of graded student writing), and resources (e.g. electronic dictionaries and thesaurus) were provided. As illustrated in Figure 2, participant recruitment and preparatory orientation were performed before the semester started. The students were familiar with the instructional and assessment plan of the course, which was a continuation of a previous course named 'Introductory English Writing'. Therefore, only a brief course introduction was provided while the teacher focused on introducing AF and PF to be utilized in the course during orientation sessions.

The intervention was implemented in alignment with the pedagogical progression of the course. During instructional weeks, the teacher focused on core writing skills and strategies. Short writing assignments were assigned during these weeks but were assessed formatively rather than collected or graded. In the writing, feedback, and revision weeks, students dedicated one session to completing their draft writing and could use one or more additional sessions to engage in dialogic written PF on the Shimo.im platform. All participants completed their revisions by the third session,
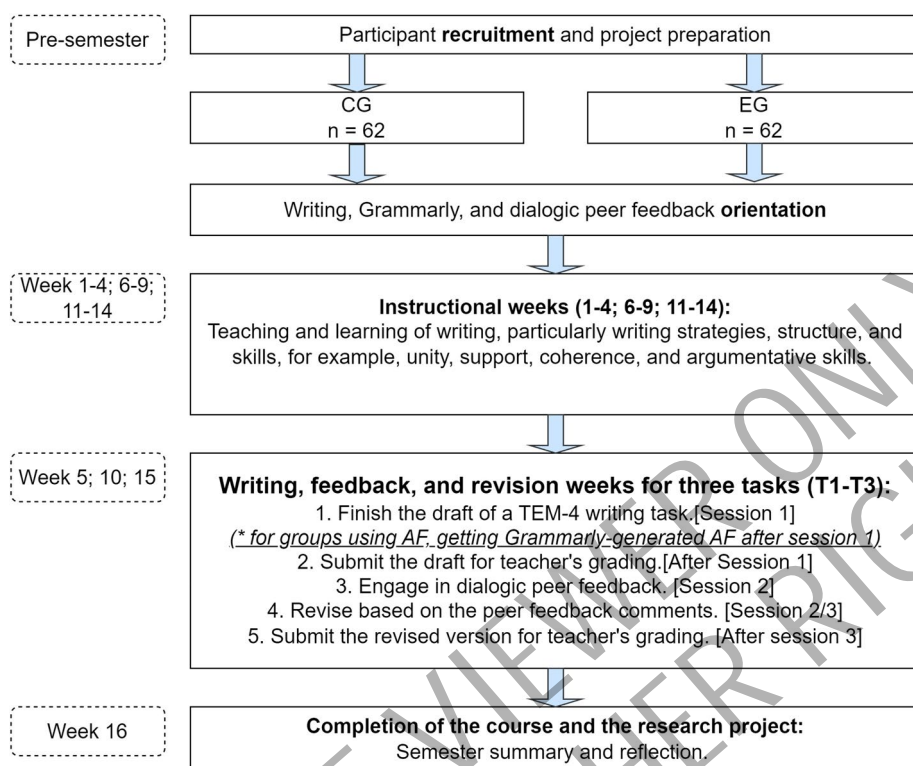
Figure 2. Research procedures.

as required. Students submitted their draft and revised writing products after the first and third sessions of the writing and revision weeks.

PF quality and writing performance improvement were assessed thrice after each during-interventional writing task (T1, T2, and T3). The teacher collected and compiled all finalized PF comments (following peer discussions and potential re-feedback from the original feedback provider in the discussion threads), along with all draft and revised writing products for further coding and analysis.

2.5. Analysis

After the coding and scoring of PF quality and performance improvement, statistical analysis was applied to examine the effects of different feedback approaches, and the developmental trajectories of feedback quality and writing performance across multiple time points in the study.

For PF quality and performance improvement, we conducted 2×3 Mixed analyses of variance (ANOVA) using R statistic software version 4.2. The intervention groups (EG and CG) were the between-subjects independent variable, and the three writing tests (T1, T2, and T3) were the within-subjects independent variable. Mauchly's test was used to check the assumption of sphericity, and Greenhouse-Geisser correction was applied in the case of non-sphericity ($p < .05$). Mixed ANOVA tests with statistically significant interaction effects were followed by post-hoc tests (i.e. Bonferroni

adjusted pairwise comparisons to control Type I errors rising from multiple comparisons) to disentangle the interaction effects over time. Effect sizes (Cohen's *d*) of post-hoc analyses on feedback quality and performance improvements were interpreted using the standards suggested by Plonsky and Oswald (2014) for L2 research.

3. Results

3.1. Feedback quality

Table 1 and Figure 3 show the descriptive statistics and developmental trends of feedback quality measurement of the two groups across the three tests. As illustrated, EG demonstrated sustained progress in the use of criteria and feedback content; however, their feedback style quality scores decreased from T1 to T3. In contrast, members in CG made progress both in feedback content and feedback style scores.

Table 2 presents the result of the 2×3 mixed ANOVA test performed on the three feedback measurements. Significant group x test interactions were observed in all three main categories of feedback quality ($ps < .001$). Group had a significant main effect on use of criteria [$F(1,122)=7.86$; $\eta^2_p =0.61$; $p < .01$]. Additionally, the variable 'test' showed significant main effects on the use of criteria [$F(1.47,179.64)=86.96$; $\eta^2_p=0.416$; $p <.001$] and feedback content [$F(1.96, 238.77)=1139.64$; $\eta^2_p=0.903$; $p <.001$].

Within-subjects post hoc comparisons reveal that: (1) only the EG learners exhibited significant improvement in use of criteria scores over time, with a large effect size [$t = 14.86$; Cohen's $d = 2.33$; $p<.001$]; (2) both CG [$t = 21.63$; $p<.001$] and EG [$t = 40.27$; $p<.001$] groups showed significant improvement in

Table 1. Descriptive statistics of feedback quality measurement across the tests.

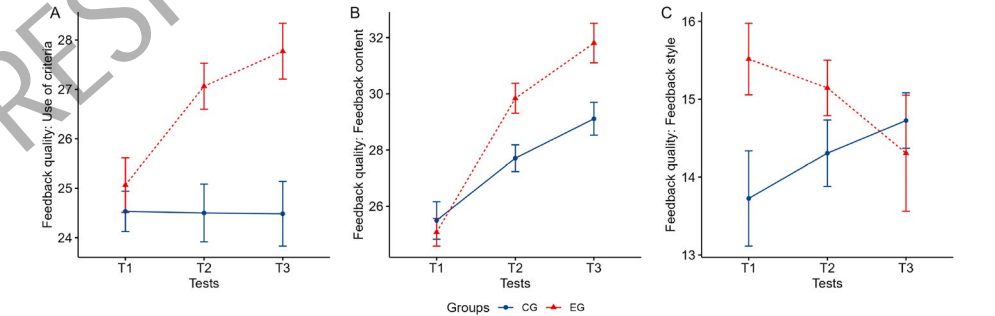| Quality features	Groups	M (SD)		
		T1	T2	T3
Use of criteria	CG	24.5 (3.22)	24.5 (4.59)	24.5 (5.15)
	EG	25.1 (4.34)	27.1 (3.66)	27.8 (4.45)
Feedback content	CG	25.5 (5.25)	27.7 (3.75)	29.1 (4.62)
	EG	25.1 (3.88)	29.8 (4.20)	31.8 (5.55)
Feedback style	CG	13.7 (4.81)	14.3 (3.36)	14.7 (2.81)
	EG	15.5 (3.62)	15.1 (2.80)	14.3 (5.87)

Figure 3. Feedback quality of the two groups across the three tests.

Table 2. Result of the mixed ANOVA on feedback quality.

Quality	Sources	df	F	η^2_p	p
Use of criteria	group	1, 122	7.86**	.061	.006
	test	1.47, 179.64	86.96***	.416	<.001
	group x test	1.47, 179.64	93.26***	.433	<.001
Feedback content	group	1, 122	3.25	.026	.074
	test	1.96, 238.77	1139.64***	.903	<.001
	group x test	1.96, 238.77	112.75***	.480	<.001
Feedback style	group	2, 185	1.12	.009	.292
	test	1.48, 180.41	0.72	.006	.447
	group x test	1.48, 180.41	20.24***	.142	<.001

Note: df: Greenhouse-Geisser corrected degrees of freedom.
*p <.05.
**p <.01.
***p <.001.

feedback content scores from T1 to T3, although the effect size of EG was larger than that of CG [Cohen's d = 5.49 vs. 2.95]; and (3) the CG group demonstrated a positive improvement [t = 3.63; p=.001] in feedback style scores over time, while the EG group experienced a significant decrease [t=-4.39; p<.001] from T1 to T3, with effect sizes indicating medium changes for both groups [Cohen's d = 0.52 and − 0.626].

Between-subjects post hoc comparisons reveal that: (1) for use of criteria, EG significantly outperformed CG with large effect sizes at T2 [t = 3.44; Cohen's d = 2.21; p<.001] and T3 [t = 3.81; Cohen's d = 2.83; p<.001]; (2) regarding feedback content, EG significantly surpassed CG with large effect sizes at T2 [t = 2.98; Cohen's d = 1.74; p<.01] and T3 [t = 2.94; Cohen's d = 2.2; p<.01]; and (3) for feedback style, significant and large-sized differences in scores were observed only at T1 [t = 2.34; Cohen's d = 0.93; p<.05].

3.2. Performance improvements

Table 3 and Figure 4 present the descriptive statistics and developmental trends of writing performance improvements of the two groups across the three tests. As illustrated, both groups made sustained improvement in content and language use along the progression of the research project. However, the CG members did not achieve greater gains in organization improvements from T1 to T3 compared to EG learners.

Table 4 presents the result of the 2×3 mixed ANOVA test performed on performance improvements. Group had a significant main effect only on content [$F(1,122)$ = 19.51; η^2_p = 0.138; p < .001]. The variable 'test' had significant main effects on content [$F(1.10, 134.62)$ = 695.92; η^2_p = 0.851; p < .001], organization [$F(1.17, 142.94)$ = 103.72; η^2_p = 0.460; p < .001], and language use [$F(1.05, 128.23)$ = 104.46; η^2_p = 0.461; p < .001]. Additionally, significant group × test interactions were found for content [$F(1.10, 134.62)$ = 63.6; η^2_p = 0.343; p < .001], organization [$F(1.17, 142.94)$ = 100.99; η^2_p = 0.453; p < .001], and language use [$F(1.05, 128.23)$ = 12.91; η^2_p = 0.096; p = .001].

Within-subjects post hoc comparisons reveal that: (1) both CG [t = 33.4, Cohen's d = 3.68, p <.001] and EG [t = 26.26, Cohen's d = 2.9, p<.001] registered significant gains in content improvements from T1 to T3, with large effect sizes; (2) a

Table 3. Descriptive statistics of writing performance improvements across the tests.

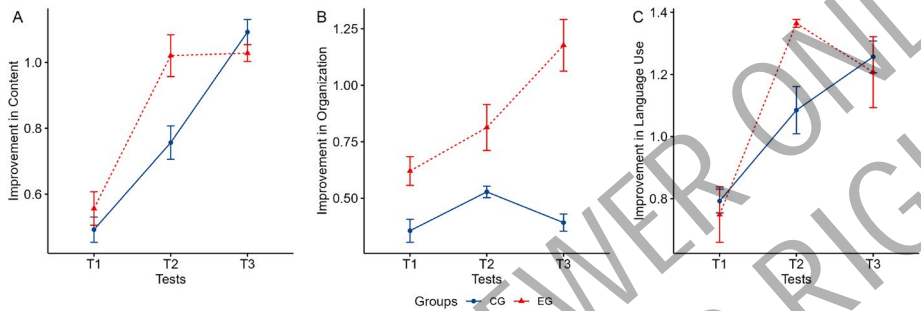| Performance dimensions	Groups	M (SD)		
		T1	T2	T3
Content	CG	0.49 (0.3)	0.76 (0.4)	1.09 (0.3)
	EG	0.56 (0.4)	1.02 (0.5)	1.03 (0.2)
Organization	CG	0.36 (0.4)	0.53 (0.2)	0.39 (0.3)
	EG	0.62 (0.5)	0.81 (0.8)	1.18 (0.9)
Language use	CG	0.79 (0.3)	1.08 (0.6)	1.26 (0.4)
	EG	0.75 (0.7)	1.36 (0.1)	1.22 (0.9)

Figure 4. Performance improvements across the three tests.

Table 4. Result of the mixed ANOVA on performance improvements.

Performance	Sources	df	F	η^2_p	p
Content	group	1, 122	1.96	.016	.493
	test	1.10, 134.62	695.92***	.851	<.001
	group x test	1.10, 134.62	63.60***	.343	<.001
Organization	group	1, 122	19.51***	.138	<.001
	test	1.17, 142.94	103.72***	.460	<.001
	group x test	1.17, 142.94	100.99***	.453	<.001
Language use	group	1, 122	0.54	.004	.464
	test	1.05, 128.23	104.46***	.461	<.001
	group x test	1.05, 128.23	12.91**	.096	.001

Note: df: Greenhouse-Geisser corrected degrees of freedom.
*p <.05.
**p <.01.
***p <.001.

significant and large difference in organization improvement scores from T1 to T3 was observed only in the EG group [t = 15.03, Cohen's d = 2.42, p<.001]; and (3) both CG [t = 23.11, Cohen's d = 1.05, p<.001] and EG [t = 26.3, Cohen's d = 1.2, p<.001] achieved significant gains in language use improvements from T1 to T3 with large effect sizes.

Between-subjects post hoc comparisons reveal that: (1) for content, EG group significantly surpassed the CG group only in T2 with a large effect size [t = 3.25; Cohen's d = 1.62; p<.01]; (2) for organization, EG group significantly outperformed the CG group with large effect sizes in T1 [t = 3.25; Cohen's d = 1.15; p<.01], T2 [t = 2.72; Cohen's d = 1.24; p<.01], and T3 [t = 6.51; Cohen's d = 3.41; p<.001]; and (3) for language use, EG group only exceeded the CG group in scores during T2 with a small effect size [t = 3.62; Cohen's d = 0.69; p<.001].

4. Discussion

4.1. Impact on feedback quality

This study found an overall positive impact of the interactive approach to integrating AF and PF. Holistically, the result could be explained from the theoretical lens of the student-feedback interaction model. According to Lipnevich and Smith (2022), successful feedback implementation relies on opportunities to internalize and generate feedback messages derived from external frames of references. This study revealed that the interactive integration of AF and PF facilitated the production of feedback comments with enhanced cognitive and constructive quality. However, the intervention appeared less effective in improving the presentation of PF messages. Therefore, we subsequently discuss the results in light of existing empirical studies by each dimension of feedback quality features.

First, the EG members following the integrative model of AF and PF integration outperformed the CG learners, who sequentially combined the two feedback sources, in producing feedback messages with better use of assessment criteria. The gains students reaped in promoting understanding of assessment criteria in feedback activities corroborate with Guo et al. (2024), who argue that the interactive approach to technology-enhanced PF enables learners to familiarize themselves with and critically achieve the assessment criteria.

Second, although both groups made significant progress over time in feedback content measures, the improvements made by EG members are more remarkable. The gains made by both groups were not unexpected as previous literature has offered sufficient evidence that the combined use of feedback from multiple sources could enrich the scope and focus of feedback messages (Niu et al., 2021; Tian & Zhou, 2020). The extra gains made by the EG members could be explained by the 'feedback → feedback-on-feedback → re-feedback' model proposed by Zhao et al. (2024): the inclusion of AF in dialogic PF expanded the chances for the dyad to engage in and substantiate the peer-based feedback discussion which would lead to the enrichment and refinement of the original feedback messages after the 're-feedback' process. In comparison, although CG learners actively participated in the dialogic interactions to generate PF, the richness of the content and the constructiveness of the comments in the produced PF messages were possibly influenced by their limited domain knowledge and feedback literacy (Latifi et al., 2023).

Third, both groups made small-sized yet statistically significant changes in the scores of feedback styles across the tests. The general weak impact of the integrative approaches to multiple feedback sources on feedback style promotion was somewhat anticipated. In previous studies, researchers have reported that the effects of different feedback interventions on the style or presentation of PF comments would be insignificant. For example, Yan (2024) found that the inclusion of students in co-creation of assessment rubrics did not result in improvement in style-related scores in feedback quality ratings. Similarly, Guo et al. (2024) noted that the AI-enhanced approach to PF enhanced the cognitive and constructive aspects of PF quality but not the affective aspect, which was a subset of what we have operationalized as feedback style in the present study.

4.2. Impact on L2 writing performance improvements

This study demonstrated that the integrative approach to integrating PF and AF prompted performance improvements across the three major dimensions of writing performance throughout the three tests. The general positive effects on writing performance were in tandem with previous literature on completing PF with AF in the context of L2 writing, where students exposed to the combined use of multiple feedback sources outperformed those who use a single form of feedback for post-draft revision (C.-C. Liu et al., 2023; Z. Zhang & Hyland, 2022).

From the dual-identity perspective of reciprocal feedback, the effects could also be explained by the theoretical insights of dialogic PF. Specifically, the engagement in PF activities enabled the feedback provider to offer more constructive and detailed comments based on their gains in understanding assessment criteria and writing standards (Gielen & De Wever, 2015); similarly, involvement in PF discussion to address confusions encountered by the feedback receivers in both improving writing and utilizing feedback could be rewarded by the providers' efforts to clarify, explain, optimize, or remake the original feedback on writing, which could possibly lead to better feedback uptake and utilization (Zhao et al., 2024).

Additionally, the study revealed that CG and EG did not have significant differences in the measures of content and language use across the three tests. The finding substantiated our understanding of the effect of feedback integration on the development of writing competence under each subcategory. In previous literature, the only study that examined such specific effects was the research by Guo et al. (2024), which identified positive effects of the integrative approach to feedback integration on language use and content measures, compared to the learners using PF without AF support. Taken together, the findings of the present study and Guo et al. (2024) suggest that integrating AF and PF could significantly impact content and language use in L2 writing, while a further shift from a complementary approach to an integrative approach did not bring additional gains.

However, Guo et al. (2024) found that the integration of AF and PF did not exert significant effects on the development of writing organization. In contrast, the present study identified statistically significant and large effects of the integrative approach of feedback integration across all three tests. The disparities in research findings could be attributed to the methodological differences between the two studies: in the Guo et al. (2024) study, the feedback provider and an AI-based tool iteratively discussed the PF messages until satisfaction, while in the present study, the assessors and assessees of the writing product engaged in iterative discussion of not only the PF messages but also the writing products. To a certain extent, Guo et al. (2024) utilized a technology-based agent to enhance only the provision of PF, which largely undermined the agency of feedback receivers in feedback discussion and application, whereas the present study enacted dialogic and technology-enhanced PF, where PF dyads actively engaged in the reciprocal and interactive processes to co-construct knowledge based on PF and AF messages (Wood, 2022; Zhao et al., 2024). Therefore, it could be concluded that the inclusion of assessees (or the authors of the writing products) in the dialogic and interactive feedback processes encourages students to address and apply feedback on global issues within their writing. From this perspective, the practice

adopted in the present study, which was similar to that in Yao et al. (2021), could be a viable solution to a key issue facing PF—the tendency to overly focus on local issues when assessing writing quality (X. Cheng et al., 2023; Niu et al., 2021).

4.3. Theoretical, pedagogical and practical implications

The study makes several theoretical contributions to our understanding of student–feedback interactions within computer-assisted language learning. Its novelty lies in its pioneering efforts to (1) identify two distinct approaches to integrating automated AF and PF in L2 writing, and (2) comparatively examine their effects on student learning. This study extends the theoretical model of student–feedback interaction (Lipnevich & Smith, 2022), particularly by exploring how multi-source external feedback can be positioned and utilized. Specifically, it broadens the role of alternative or technological feedback sources—not merely as supplementary information to support peer-based scaffolding, but also as components to form a *combination of comparators* (see Nicol, 2021) that open up opportunities for learners to internalize external feedback and transform it into self-feedback during the processes of (collaborative) feedback generation. Furthermore, the interactional approach to integrating feedback enriches our current procedural understanding of PF. In existing models of dialogic PF (e.g. Zhao et al., 2024), the 'feedback-on-feedback' phase is framed as reciprocal interactions to clarify ambiguities or inaccuracies in PF comments. This study expands the scope of this phase from the traditional notion of 'PF on PF' to include 'PF on multi-source feedback' (as in the current study and Yao et al., 2021), and even 'multi-source feedback on PF' (as in Guo et al., 2024), by situating PF within an interactive and technology-enhanced environment. The study thus contributes to the growing understanding of how the interactional integration of technology can enhance assessment practice and learning outcomes of collaborative learning environments.

From a pedagogical perspective, the study had several implications for researchers and practitioners of contemporary assessment and feedback in L2 writing classrooms. First, the study revealed that an interactive approach to technology-enhanced PF remarkably enhanced the quality of student-generated feedback messages and learning outcomes. However, to date, technology- and human-based sources of feedback are not effectively and systematically integrated in L2 writing classrooms, compared to the substantial body of literature on their individual implication or the comparison of effects between them. Noticeably, in recent years, more powerful and intelligent technology-based tools capable of simulating human evaluative judgement, such as the AI-based EvaluMate system in Guo et al. (2024), exhibited greater potential for educators to interactively apply AF to complement the use of formative PF. Thus, in authentic L2 writing environments, the interactive integration of multiple feedback sources should be encouraged and implemented for sustained development of students' writing competence and feedback literacy. Second, the result that emerged from this experimental study revealed that the change in the design of the feedback intervention would elicit rather remarkable effects on feedback and learning. Thus, in the contemporary era, the role played by L2 teachers as researchers and designers of feedback interventions should not be downplayed or neglected. The focus on learner centeredness and agency in the *new feedback paradigm* should not necessarily

overshadow the visibility and significance of teacher-as-researchers (Jensen et al., 2023). On the contrary, it is the educators' duty to design learning environments facilitating opportunities for students to engage in and make good use of the feedback processes, ultimately leading to self-feedback generation and feedback uptake (Carless, 2022). Following this line of reasoning, the current study suggested that, to make PF realize its potential, efforts made by teachers and educators to finetune the prototypical designs towards an interactive and self-regulatory approach to feedback are highly valued and expected (Evans et al., 2024; Yan et al., 2025).

It should be noted that the current study also has significant implications for designing and implementing PF interventions. Based on the re-conceptualization of first- and second-order scaffolding for feedback research in Noroozi et al. (2018), it is evident that most existing research on the complementary integration of AF and PF falls under first-order scaffolding. These studies typically offer students extended sources of feedback during specific tasks, while students remain passive recipients of 'feedback-as-information' rather than active agents engaged in 'feedback-as-process' or related interactions (Winstone et al., 2022). In contrast, the interactional approach to feedback integration could be understood as second-order scaffolding with their focus on enabling learners to develop domain knowledge (L2 writing competence) and feedback outcomes through enhanced interactivity. Drawing on the current study, the core principle of second-order scaffolding is to create opportunities for feedback interactions in which students can effectively generate and utilize self-feedback based on their engagement in interactively using and explicitly comparing feedback from multiple sources (Lipnevich & Smith, 2022; Nicol, 2021). Specifically, in designing the PF innovation of this study, we prioritized learner collaboration by embedding technological support as an integral component to facilitate the progression and outcomes of peer-based interaction.

Looking retrospectively into the literature, while it is widely acknowledged that feedback activity design plays a crucial role in determining its effects (Kerman et al., 2024), most existing studies have predominantly employed first-order scaffolding elements, such as task-specific supportive materials, anonymity policies, and pre-feedback training, in their feedback intervention designs (Bürgermeister et al., 2021). The current study highlights that incorporating second-order scaffolding elements—specifically changes in the modes of interactions among students and between students and feedback—can lead to significant and longitudinal improvements in learning outcomes.

5. Limitations and future research directions

The study might be one of the earliest scholarly efforts to compare the effects of the interactive vs. complementary approaches to integrating AF and PF in the context of L2 writing. Naturally, it faces limitations that could be addressed by future research. First, the study adopted a short-term longitudinal and quantitative approach to examine the effects of the two integrative approaches. As a result, it does not illuminate the nuances of students' engagement with the dialogic feedback processes or the characteristics of the feedback messages that emerged from peer discussions. In particular, the interactive approach to AF-PF integration may have negative impacts on L2 writers' learning experiences, for example, it could be overly time-consuming and increase

cognitive load. We recommend future studies to conduct mixed-method or naturalistic research on students' perceptions and behavioral patterns to gain a more comprehensive understanding of feedback interactions.

Second, while the study focused on the cognitive aspects of feedback, it did not deeply address the emotional and relational dimensions of feedback processes, which are increasingly recognized as critical in the PF literature. Although previous literature informed that the integration of multiple sources of feedback promotes rather than impedes learner mindset and motivation (Yao et al., 2021), students might still feel overwhelmed by the additional cognitive load, time expenditure, and extended conversational interactions with both peers and technology-based agents. Future studies could explore how emotional engagement, trust between peers, and the development of feedback relationships impact the effectiveness of AF and PF integration for a more holistic understanding.

Finally, the study has broadly classified existing approaches to integrating AF and PF into two major categories. While the research findings elucidated the general benefits of the interactive approach compared to the complementary approach, future research could continue to explore and compare the effects of different feedback interventions with detailed methodological considerations within the interactive category. For example, Guo et al. (2024) applied an technology-enhanced PF process composing of an AI system and feedback provider, while the current study and Yao et al. (2021) included both PF provider and recipient in discussing AF messages. Future researchers could refine intervention design specifications to identify best practices for integrating PF with AF or intelligent agents.

Acknowledgement

Disclosure statement

Funding

Notes on contributors

Dr. *Da 'Alex' Yan* has research interests in computer-assisted language learning, formative assessment, and human-(Gen)AI interactions. He has published papers in renowned international journals such as *Assessment & Evaluation in Higher Education*, *Language Learning & Technology*, *Education and Information Technologies*, *Interactive Learning Environments*, *Kybernetes,* and *Humanities and Social Sciences Communications*.

Yu Gao is a lecture at Xinyang Agriculture and Forestry University, China. Her research interests include translation studies, applied linguistics, and computer-assisted language learning.

ORCID

Da "Alex" Yan 🄳 http://orcid.org/ 0000-0002-1265-9772

References

Aben, J. E. J., Timmermans, A. C., Dingyloudi, F., Lara, M. M., & Strijbos, J.-W. (2022). What influences students' peer-feedback uptake? Relations between error tolerance, feedback tolerance, writing self-efficacy, perceived language skills and peer-feedback processing. *Learning and Individual Differences*, *97*, 102175. https://doi.org/10.1016/j.lindif.2022.102175

Bürgermeister, A., Glogger-Frey, I., & Saalbach, H. (2021). Supporting peer feedback on learning strategies: Effects on self-efficacy and feedback quality. *Psychology Learning & Teaching*, *20*(3), 383–404. https://doi.org/10.1177/14757257211016604

Cao, Z., Yu, S., & Huang, J. (2019). A qualitative inquiry into undergraduates' learning from giving and receiving peer feedback in L2 writing: Insights from a case study. *Studies in Educational Evaluation*, *63*, 102–112. https://doi.org/10.1016/j.stueduc.2019.08.001

Carless, D. (2022). From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes. *Active Learning in Higher Education*, *23*(2), 143–153. https://doi.org/10.1177/1469787420945845

Chan, S., Lo, N., & Wong, A. (2025). Leveraging generative AI for enhancing university-level english writing: Comparative insights on automated feedback and student engagement. *Cogent Education*, https://www.tandfonline.com/doi/abs/10.1080/2331186X.2024.2440182 https://doi.org/10.1080/2331186X.2024.2440182

Chen, H., & Hu, G. (2025). Investigating relationships among students' affective, behavioral and cognitive engagement with peer feedback on EFL writing. *Studies in Educational Evaluation*, *84*, 101430. https://doi.org/10.1016/j.stueduc.2024.101430

Cheng, L., Li, Y., Su, Y., & Gao, L. (2023). Effect of regulation scripts for dialogic peer assessment on feedback quality, critical thinking and climate of trust. *Assessment & Evaluation in Higher Education*, *48*(4), 451–463. https://doi.org/10.1080/02602938.2022.2092068

Cheng, X., Liu, Y., & Wang, C. (2023). Understanding student engagement with teacher and peer feedback in L2 writing. *System*, *119*, 103176. https://doi.org/10.1016/j.system.2023.103176

Cheong, C. M., Luo, N., Zhu, X., Lu, Q., & Wei, W. (2023). Self-assessment complements peer assessment for undergraduate students in an academic writing task. *Assessment & Evaluation in Higher Education*, *48*(1), 135–148. https://doi.org/10.1080/02602938.2022.2069225

Dao, P., Duong, P.-T., & Nguyen, M. X. N. C. (2023). Effects of SCMC mode and learner familiarity on peer feedback in L2 interaction. *Computer Assisted Language Learning*, *36*(7), 1206–1235. https://doi.org/10.1080/09588221.2021.1976212

Elola, I., & Oskoz, A. (2016). Supporting second language writing using multimodal feedback. *Foreign Language Annals*, *49*(1), 58–74. https://doi.org/10.1111/flan.12183

Er, E., Dimitriadis, Y., & Gašević, D. (2021). Collaborative peer feedback and learning analytics: Theory-oriented design for supporting class-wide interventions. *Assessment & Evaluation in Higher Education*, *46*(2), 169–190. https://doi.org/10.1080/02602938.2020.1764490

Evans, C., Kay, W., Amici-Dargan, S., González, R. D. M., Donert, K., & Rutherford, S. (2024). Developing a scale to explore self-regulatory approaches to assessment and feedback with academics in higher education. *Frontiers in Psychology*, *15*, 1357939. https://doi.org/10.3389/fpsyg.2024.1357939

Fan, N., & Ma, Y. (2022). The effects of automated writing evaluation (AWE) feedback on students' english writing quality: A systematic literature review. *Language Teaching Research Quarterly*, *28*, 53–73. https://doi.org/10.32038/ltrq.2022.28.03

García, E. C., Ferré, A. J., Molins, L. L., & Gironelli, L. M. (2024). Improvement in the quality of feedback as an indication of the development of evaluative judgement. *Assessment & Evaluation in Higher Education*, 49(6), 824–837. https://doi.org/10.1080/02602938.2024.2329170

Gielen, M., & De Wever, B. (2015). Structuring the peer assessment process: A multilevel approach for the impact on product improvement and peer feedback quality. *Journal of Computer Assisted Learning*, 31(5), 435–449. https://doi.org/10.1111/jcal.12096

Golparvar, S. E., & Abolhasani, H. (2022). Unpacking the contribution of linguistic features to graph writing quality: An analytic scoring approach. *Assessing Writing*, 53, 100644. https://doi.org/10.1016/j.asw.2022.100644

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. *Internet and Higher Education*, 63, 100962. https://doi.org/10.1016/j.iheduc.2024.100962

Haro, V. A., Noroozi, O., Biemans, H. J. A., Mulder, M., & Banihashem, S. K. (2023). How does the type of online peer feedback influence feedback quality, argumentative essay writing quality, and domain-specific learning? *Interactive Learning Environments*, 32(9), 5459–5478. https://doi.org/10.1080/10494820.2023.2215822

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. *Computers & Education*, 71, 133–152. https://doi.org/10.1016/j.compedu.2013.09.019

Hyland, K., & Hyland, F. (2019). Contexts and Issues in Feedback on L2 Writing. In F. Hyland (Ed.), *Feedback in second language writing: Contexts and issues* (2nd ed., pp. 1–22). Cambridge University Press. https://doi.org/10.1017/9781108635547.003

Ibarra-Sáiz, M. S., Rodríguez-Gómez, G., & Boud, D. (2020). Developing student competence through peer assessment: The role of feedback, self-regulation and evaluative judgement. *Higher Education*, 80(1), 137–156. https://doi.org/10.1007/s10734-019-00469-2

Jensen, L. X., Bearman, M., & Boud, D. (2023). Feedback encounters: Towards a framework for analysing and understanding feedback processes. *Assessment & Evaluation in Higher Education*, 48(1), 121–134. https://doi.org/10.1080/02602938.2022.2059446

Kasprowicz, R., & Marsden, E. (2018). Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form-meaning practice. *Applied Linguistics*, 39(6), 886–911. https://doi.org/10.1093/applin/amw051

Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies*, 29(1), 763–813. https://doi.org/10.1007/s10639-023-12273-8

Kerman, N. T., Noroozi, O., Banihashem, S. K., Karami, M., & Biemans, H. J. A. (2024). Online peer feedback patterns of success and failure in argumentative essay writing. *Interactive Learning Environments*, 32(2), 614–626. https://doi.org/10.1080/10494820.2022.2093914

Latifi, S., Noroozi, O., & Talaee, E. (2023). Worked example or scripting? Fostering students' online argumentative peer feedback, essay writing and learning. *Interactive Learning Environments*, 31(2), 655–669. https://doi.org/10.1080/10494820.2020.1799032

Li, X., & Ke, P. (2025). How L2 student writers engage with automated feedback: A longitudinal perspective. *Assessing Writing*, 64, 100919. https://doi.org/10.1016/j.asw.2025.100919

Lipnevich, A. A., & Smith, J. K. (2022). Student – feedback interaction model: Revised. *Studies in Educational Evaluation*, 75, 101208. https://doi.org/10.1016/j.stueduc.2022.101208

Liu, C.-C., Liu, S.-J., Hwang, G.-J., Tu, Y.-F., Wang, Y., & Wang, N. (2023). Engaging EFL students' critical thinking tendency and in-depth reflection in technology-based writing contexts: A peer assessment-incorporated automatic evaluation approach. *Education and Information Technologies*, 28(10), 13027–13052. https://doi.org/10.1007/s10639-023-11697-6

Liu, W. (2024). A systematic review of automated writing evaluation feedback: Validity, effects and students' engagement. *Language Teaching Research Quarterly*, 45, 86–105. https://doi.org/10.32038/ltrq.2024.45.05

Loncar, M., Schams, W., & Liang, J.-S. (2023). Multiple technologies, multiple sources: Trends and analyses of the literature on technology-mediated feedback for L2 english writing

published from 2015-2019. *Computer Assisted Language Learning*, *36*(4), 722–784. https://doi.org/10.1080/09588221.2021.1943452

Ma, M. (2023). Using teacher-supported peer feedback to develop student feedback literacy in L2 writing: An instructional model. *Feedback Research in Second Language*, *1*, 76–92. https://doi.org/10.32038/frsl.2023.01.05

Man, D., Kong, B., & Chau, M. H. (2022). Developing student feedback literacy through peer review training. *RELC Journal*, *55*(2), 408–421. https://doi.org/10.1177/00336882221078380

Myers, T., & Buchanan, J. (2025). Dialogism in feedback literacies: A critical review. *Assessment & Evaluation in Higher Education*, *50*(6), 846–860. https://doi.org/10.1080/02602938.2025.2478159

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, *46*(5), 756–778. https://doi.org/10.1080/02602938.2020.1823314

Niu, R., Shan, P., & You, X. (2021). Complementation of multiple sources of feedback in EFL learners' writing. *Assessing Writing*, *49*, 100549. https://doi.org/10.1016/j.asw.2021.100549

Noroozi, O., Alqassab, M., Taghizadeh Kerman, N., Banihashem, S. K., & Panadero, E. (2025). Does perception mean learning? Insights from an online peer feedback setting. *Assessment & Evaluation in Higher Education*, *50*(1), 83–97. https://doi.org/10.1080/02602938.2024.2345669

Noroozi, O., Kirschner, P. A., Biemans, H. J. A., & Mulder, M. (2018). Promoting argumentation competence: Extending from first- to second-order scaffolding through adaptive fading. *Educational Psychology Review*, *30*(1), 153–176. https://doi.org/10.1007/s10648-017-9400-z

Öztürk, N. B., Şahin, M. G., & İlhan, M. (2019). An analysis of scoring via analytic rubric and general impression in peer assessment. *Turkish Journal of Education*, *8*(4), 258–275. https://doi.org/10.19128/turje.609073

Panadero, E., Jonsson, A., & Alqassab, M. (2018). Providing formative peer feedback: What do we know?. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge Handbook of Instructional Feedback.* (pp. 409–431) Cambridge University Press. https://doi.org/10.1017/9781316832134.020

Patchan, M. M., Schunn, C. D., & Correnti, R. J. (2016). The nature of feedback: How peer feedback features affect students' implementation rate and quality of revisions. *Journal of Educational Psychology*, *108*(8), 1098–1120. https://doi.org/10.1037/edu0000103

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Qian, J., & Li, D. (2025). Toward a better understanding of student engagement with peer feedback: A longitudinal study. *International Review of Applied Linguistics in Language Teaching*, *63*(1), 709–734. https://doi.org/10.1515/iral-2023-0108

Shao, J., Cheng, L., Wang, Y., Li, K., & Li, Y. (2023). How peer feedback with regulation scripts contributes to the development of critical thinking in dialogues: Strengthening cognitive and affective feedback content. *Interactive Learning Environments*, *32*(10), 6121–6140. https://doi.org/10.1080/10494820.2023.2251040

Sun, H., & Wang, M. (2025). Effects of teacher intervention and type of peer feedback on student writing revision. *Language Teaching Research*, *29*(3), 1094–1113. https://doi.org/10.1177/13621688221080507

Tan, S., Cho, Y. W., & Xu, W. (2023). Exploring the effects of automated written corrective feedback, computer-mediated peer feedback and their combination mode on EFL learner's writing performance. *Interactive Learning Environments*, *31*(10), 7276–7286. https://doi.org/10.1080/10494820.2022.2066137

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

Van Blankenstein, F. M., Dirkx, K. J. H., & De Bruycker, N. M. F. (2025). Ask your peer! How requests for peer feedback affect peer feedback responses. *Educational Research and Evaluation*, *30*(1–2), 36–57. https://doi.org/10.1080/13803611.2024.2376832

Van der Kleij, F. M. (2024). Supporting students to generate self-feedback: Critical reflections on the special issue. *Studies in Educational Evaluation*, 81, 101348. https://doi.org/10.1016/j.stueduc.2024.101348

Vuogan, A., & Li, S. (2023). Examining the effectiveness of peer feedback in second language writing: A meta-analysis. *Tesol Quarterly*, 57(4), 1115–1138. https://doi.org/10.1002/tesq.3178

Winstone, N., Boud, D., Dawson, P., & Heron, M. (2022). From feedback-as-information to feedback-as-process: A linguistic analysis of the feedback literature. *Assessment & Evaluation in Higher Education*, 47(2), 213–230. https://doi.org/10.1080/02602938.2021.1902467

Wood, J. (2022). Making peer feedback work: The contribution of technology-mediated dialogic peer feedback to feedback uptake and literacy. *Assessment & Evaluation in Higher Education*, 47(3), 327–346. https://doi.org/10.1080/02602938.2021.1914544

Wu, W., Huang, J., Han, C., & Zhang, J. (2022). Evaluating peer feedback as a reliable and valid complementary aid to teacher feedback in EFL writing classrooms: A feedback giver perspective. *Studies in Educational Evaluation*, 73, 101140. https://doi.org/10.1016/j.stueduc.2022.101140

Xu, L., & Zhang, T. (2023). Engaging with multiple sources of feedback in academic writing: Postgraduate students' perspectives. *Assessment & Evaluation in Higher Education*, 48(7), 995–1008. https://doi.org/10.1080/02602938.2022.2161089

Yan, D. (2024). Rubric co-creation to promote quality, interactivity and uptake of peer feedback. *Assessment & Evaluation in Higher Education*, 49(8), 1017–1034. https://doi.org/10.1080/02602938.2024.2333005

Yan, D. "Alex," Jia, C., & Tian, F. (2025). Automated feedback complements online dialogic peer feedback: Fostering L2 writers' feedback quality and self-regulatory writing strategy use. *Interactive Learning Environments*, 1–15. https://doi.org/10.1080/10494820.2025.2491620

Yang, C. T.-Y., & Chen, H. H.-J. (2025). ChatGPT and L2 chinese writing: Evaluating the impact of model version and prompt language on automated corrective feedback. *Computer Assisted Language Learning*, https://www.tandfonline.com/doi/abs/10.1080/09588221.2025.2453205 https://doi.org/10.1080/09588221.2025.2453205

Yao, Y., Wang, W., & Yang, X. (2021). Perceptions of the inclusion of Automatic Writing Evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing*, 50, 100568. https://doi.org/10.1016/j.asw.2021.100568

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, 49(4), 461–493. https://doi.org/10.1017/S0261444816000161

Zhai, N., & Ma, X. (2023). The Effectiveness of Automated Writing Evaluation on Writing Quality: A Meta-Analysis. *Journal of Educational Computing Research*, 61(4), 875–900. https://doi.org/10.1177/07356331221127300

Zhang, X., & Lu, X. (2022). Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: The case of two genres. *Assessing Writing*, 51, 100597. https://doi.org/10.1016/j.asw.2021.100597

Zhang, Y., & Gao, Y. (2024). Exploring the dynamics of student engagement with receiving peer feedback in L2 writing. *Assessing Writing*, 60, 100842. https://doi.org/10.1016/j.asw.2024.100842

Zhang, Z., & Hyland, K. (2022). Fostering student engagement with feedback: An integrated approach. *Assessing Writing*, 51, 100586. https://doi.org/10.1016/j.asw.2021.100586

Zhao, Y., Zhang, F., Schunn, C. D., He, P., Li, D., & Zhao, Y. (2024). Feedback, feedback-on-feedback and re-feedback: Effects of written dialogic peer feedback on English as a foreign language writing. *Assessment & Evaluation in Higher Education*, 49(4), 523–538. https://doi.org/10.1080/02602938.2023.2278017

Zong, Z., Schunn, C. D., & Wang, Y. (2021). Learning to improve the quality peer feedback through experience with peer feedback. *Assessment & Evaluation in Higher Education*, 46(6), 973–992. https://doi.org/10.1080/02602938.2020.1833179

Appendix A

Table A1. Feedback quality scoring rubric.

Category	Subcategory	Good feedback		Average feedback		Poor feedback	
Use of criteria	Coverage	Comments on all feedback aspects, in combination with the associated criteria	20	Comments on some feedback aspects, in combination with the associated criteria	10	None or minimal comments	0
	Clarification	Clarification of all comments on feedback aspects against the associated criteria	15	Clarification of some comments on feedback aspects against the associated criteria	7	None or minimal clarification of comments on feedback aspects	0
Feedback contents	Problem Identification	Included explicit and localized identification of problem	10	Included identification of problem without localization of identified problem	5	None or minimal explicit identification of problem	0
	Problem Justification	Included elaborations and justifications of identified problem	10	Included elaborations but not justifications of identified problem	5	None or minimal elaborations and justifications of identified problem	0
	Reflection stimulation	Multiple questions which stimulate reflection	10	Few question which stimulates reflection	5	No asked questions to stimulate reflection	0
	Constructiveness	Included recommendations and action plans for further improvements	10	Included recommendations but not action plans for further improvements	5	None or minimal recommendations or action plans for further improvements	0
Feedback styles	Structure	Clear structure	10	Unclear structure	5	No structure	0
	Valence	Equilibrium between positive and negative comments	10	Occasionally have positive comments	5	Mainly negative comments	0
	Affective/Tone	Included encouraging and positive emotions	5	Did not include either negative or positive emotions	2	Included discouraging and negative emotions	0

Note: Total score = 100 (35 + 40 + 25); adapted from Gielen and De Wever (2015) and Kerman et al. (2024).