Automated feedback complements online dialogic peer feedback: Fostering L2 writers' feedback quality and self-regulatory writing strategy use

Da "Alex" Yan[a, *], Chenjin Jia[b], and Feng Tian[c,d]

[a]*School of Foreign Languages, Xinyang Agriculture and Forestry University, Xinyang, China.*

[b]*School of Literature and Journalism, Yibin University, Yibin, China.*

[c]*Center of Instructional Technology and Multimedia, University of Science, Malaysia, Penang, Malaysia.*

[d]*School of Foreign Languages, Ningxia University, Yinchuan, China.*

* Correspondence: [Postal address as affiliation]; email: alexyan1987@outlook.com

This study examined the effects of the complementary integration of automated feedback provided by automated writing evaluation (AWE) systems and online dialogic peer feedback on the quality of feedback and the use of self-regulatory writing strategies among second language (L2) writers. The participants were 68 Chinese L2 students who were assigned to two experimental conditions: one using automated feedback to complement dialogic peer feedback, and the other relying solely on dialogic peer feedback. The participants engaged in dyadic peer feedback for three writing tasks during a 14-week writing course. After collecting peer feedback comments and administering the self-regulatory writing strategy questionnaire, the researchers coded and statistically analyzed the data. Results indicate that integrating automated and peer feedback enhances the cognitive and constructive quality of peer feedback messages; however, there was no significant difference in the affective features of feedback between the two groups. Additionally, students in the integrated condition more frequently employed self-regulatory writing strategies during the text generation, self-monitoring, and revision phases, while their strategy use during the planning phase was comparable to that of students in the peer feedback-only condition. These findings have implications for educators and researchers aiming to advance peer feedback practices in L2 writing classrooms.

Keywords: online dialogic peer feedback, automated feedback, L2 writing, feedback quality, self-regulated learning

Introduction

Peer feedback has been widely implemented in second language (L2) writing classrooms as an instructional scaffold to support the development of writing competence (Lee, 2017). Recently, peer feedback has increasingly been integrated into online settings to address teachers' high workload and time constraints when providing feedback in large classrooms (Er et al., 2021). While the efficacy of peer feedback in fostering writing development is widely recognized (Vuogan & Li, 2023), researchers have also identified challenges that students encounter when attempting to utilize peer feedback effectively. For instance, students, particularly those who struggle to provide constructive feedback without additional guidance and support, experienced significant cognitive barriers in peer feedback processes (Cheng et al., 2023; Kerman et al., 2024). Consequently, a growing body of research has explored various forms of support and scaffolding to enhance both the quality of peer-generated feedback and learners' engagement with and use of feedback messages (Noroozi et al., 2023; Author, 2024).

A supportive measure adopted by researchers and educators in the context of L2 writing is the application of automated feedback to complement peer feedback (Loncar et al., 2023; Tan et al., 2023). A review of related literature shows that such an approach to integrating feedback sources has proven effective in promoting learner mindset and motivation (Yao et al., 2021), improving writing performance (Kao & Reynolds, 2024), reducing anxiety (Liu et al., 2023), and enhancing feedback engagement (Zhang & Hyland, 2022). However, further investigation into the effects of integrating automated and peer feedback is warranted. First, existing studies offer limited insights into the effects of the integration of feedback on the improvement of peer feedback quality. In

response to Panadero's (2023) call that "feedback can be … intervention [or] … dependent variable" in research (p. 198), a closer examination of feedback quality could contribute to a more comprehensive evaluation of the integrative approach as both an intervention and an instructional scaffold. The importance of feedback quality cannot be overstated, as it is a prerequisite for active feedback uptake and successful, feedback-informed revisions (Er et al., 2021; Zong et al., 2021). Second, the current body of scholarship has largely neglected the effects of feedback integration on the development of higher-order capabilities, such as self-regulated learning (SRL). To achieve long-term effects of feedback on performance and learning, feedback messages must first be internalized and transformed into self-feedback through a process significantly influenced by students' self-regulatory abilities (Nicol & Macfarlane-Dick, 2006). Unfortunately, in the context of integrating automated and peer feedback in L2 writing classrooms, little is known about how students might strategically regulate this process and its subsequent impact on learning.

Against the above-stated backdrops, the current study examined the effects of the complementary integration of automated and peer feedback on the quality of peer feedback messages and the use of self-regulatory writing strategies. Following current understanding of effective feedback practices, for instance, utilization of student-centered (Reynolds & Zhang, 2023) and technology-enhanced feedback (Kao & Reynolds, 2024), this study was conducted in a 14-week L2 writing course at a Chinese university. This study contributes to the literature on peer feedback and L2 writing and offers insights for researchers and educators interested in advancing contemporary formative feedback practices in L2 writing.

Literature review

Online dialogic peer feedback

The understanding of peer feedback continues to evolve. Peer feedback is defined as a reciprocal process whereby students provide and receive evaluative comments on their work or performance (Hattie & Timperley, 2007; Winstone et al., 2022). In recent years, various online learning environments have been implemented to support the peer feedback process and its impact on writing pedagogy (Latifi et al., 2021; Noroozi et al., 2023). Compared to in-person peer feedback, online peer feedback leverages the advantages of reciprocal and anonymous performance evaluation, as well as a relatively continuous workflow (Noroozi et al., 2023).

At the same time, researchers in the field of assessment and feedback have highlighted the added value of dialogic peer feedback, a collaborative meaning-making process in which students actively engage in dialog with peers to co-construct feedback and knowledge (Er et al., 2021). Wood (2022) argues that technology-mediated dialogic peer feedback enables students to collaboratively create, process, and utilize actionable feedback, leading to improved feedback literacy and uptake. Zhao et al. (2024) proposed a three-step model to explain the mechanism of dialogic peer feedback. The model's second phase, termed "feedback-on-feedback," refers to the discussion between feedback receivers and providers following the initial feedback exchange. This phase is followed by a third stage of "re-feedback," in which the original feedback providers refine or adjust their comments based on the outcomes of the feedback discussion.

In summary, online dialogic peer feedback has been contended as a technology-mediated and interactional process of evaluation and knowledge co-construction. Therefore, the current study drew upon the above understandings and aimed to examine the effects of a feedback intervention in a dialogic online peer feedback setting. The

subsequent sections review two aspects of successful peer feedback practices: quality of and self-regulatory learning in peer feedback.

Peer feedback quality

A central challenge that hinders the effectiveness of peer feedback is its unsatisfactory quality. To begin with, peer feedback comments produced by students are frequently criticized for being *cognitively* superficial due to their limited domain knowledge (Latifi et al., 2023). In the realm of L2 writing, a widely recognized concern regarding peer feedback quality is its disproportionate focus on local issues within writing products (Cheng et al., 2023; Tan et al., 2024).

Moreover, the quality challenges faced by peer feedback also have an *affective* dimension. Receivers' skepticism toward feedback stems from their distrust in peers' knowledge as well as environmental factors such as an inharmonious classroom atmosphere (Ibarra-Sáiz et al., 2020). Additionally, peer feedback has been found to lack *constructive* solutions and actionable suggestions for further improvements (Foo, 2021). Previous literature has attributed this shortcoming to students' limited higher-order thinking skills and evaluative judgment capabilities (Latifi & Noroozi, 2021). Taken together, these quality issues reduce the cognitive and affective engagement with peer feedback, thereby diminishing its impact on learning improvements (Cheng et al., 2014).

To enhance peer feedback quality, existing studies have employed either cognitive or social scaffolds. The former enables students to co-construct meaningful evaluative judgments while the latter promotes deeper engagement with feedback processes (Bürgermeister et al., 2021). In the present study, we designed a feedback intervention by integrating automated feedback with online dialogic peer feedback in a complementary fashion. Following the aforementioned categorization, the intervention

exerts a double-faceted effect on writing development. First, it serves as a cognitive scaffold, expanding the scope of available feedback comments through the integration of multiple feedback sources. Second, it functions as a communicative facilitator, leveraging technology-based platforms to encourage dialogic interaction during peer feedback encounters.

Peer feedback and self-regulated learning in L2 writing context

Beyond its impact on learning/performance and motivation/affect, feedback serves a third purpose: supporting self-regulated learning (SRL), which refers to students' ability to regulate their (meta)cognition, behavior, emotions, and motivation during learning processes (Panadero & Lipnevich, 2022). Defined as a process encompassing forethought, performance, and self-reflection, SRL is considered a key factor that comprehensively influences cognitive processes in writing development (Teng & Zhang, 2020).

Effective self-regulation of the writing process requires L2 writers to proactively deploy strategies to achieve the learning goals by activating, sustaining, and adjusting the cognitive, affective, and behavioral aspects of learning (Zhang & Zhang, 2024). Since both SRL and the writing process are multi-phase constructs, existing SRL models can be integrated into a framework aligned with the process-genre approach to L2 writing (Shen & Wang, 2024). Based on the integrated framework, we can identify the following four major types of self-regulatory writing strategies: (1) before writing, learners employ *planning* strategies to analyze task requirements, set goals, and generate ideas; (2) during writing, they retrieve relevant information from long-term memory, translate ideas into textual output through *text-generating* strategies; (3) learners also engage in *self-monitoring* to refine their work against initial plans during

writing; and (4) after writing, *revising* strategies are activated to identify and address problems in writing products (Bai et al., 2022).

Formative feedback is linked to SRL due to its iterative nature and ability to elicit monitoring and self-reflection in learning (Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006). Scientific literature indicates that teacher feedback influences EFL writers' self-regulatory writing strategy use, especially their abilities to set goals, metacognitively monitor learning processes, process text, and handle learning resources (Yang et al., 2022, 2023). Similarly, portfolio feedback is also believed to be an effective practice for fostering L2 writers' critical awareness in writing evaluation (Lam, 2015; Mak & Wong, 2018). Noticeably, researchers have highlighted the potential of self-assessment and peer feedback to effectively scaffold SRL in classroom settings (Panadero et al., 2016).

Compared to other feedback forms, dialogic peer feedback has the following advantages in supporting SRL. First, its reciprocal nature can stimulate feedback recipients' SRL development while simultaneously co-regulating feedback providers' writing skills and evaluative judgment (Alemdag & Yildirim, 2022). Second, the dialogic discussion on feedback, also known as back-evaluation, fosters sustained, in-depth, and critical reflection (Misiejuk & Wasson, 2021).

Unfortunately, the use of dialogic peer feedback as an intervention to promote SRL development was scarcely found (Yang et al., 2023). Moreover, following the accounts on internal feedback generation (Lipnevich & Smith, 2022; Nicol & Macfarlane-Dick, 2006), effective translating from external feedback sources into learning interactions could substantially enhance learners' SRL capabilities. Therefore, the complementary use of both automated and peer feedback to foster L2 writers' self-regulation warrants further investigation and offers valuable insights for the existing

literature. In the following subsection, empirical evidence on integrating automated and peer feedback is synthesized.

Automated feedback complementing peer feedback: empirical evidence

A few studies have examined the effects of combining automated and peer feedback in L2 writing settings. Zhang and Hyland (2022) conducted a case study in a writing class that enacted an integrated approach to combining feedback from automated writing evaluation (AWE) systems, peers, and the teacher. The study showed that the combination of multiple feedback sources enhanced students' cognitive, affective, and behavioral engagement with feedback and promoted effective revision. Liu et al. (2023) consolidated and extended the above observation through an experimental study. The researchers contended that the peer-automated-combined-feedback intervention enables learners to outperform in writing performance, motivation, critical thinking, and anxiety reduction. Similarly, the factorial experimental study by Tan et al. (2023) also discovered that the combination of AWE and asynchronous peer feedback facilitated learners' gains in quality of writing products.

However, only the study by Yao et al. (2021) implemented the complementary integration of AWE and dialogic peer feedback, engaging participants in discussion about peers' writing products and Pigai-generated corrective feedback comments before asking the students to carry out revision based on the integrated feedback messages. The short-term longitudinal study revealed that the integration of automated and dialogic peer feedback promotes language mindset and motivation.

In summary, while the complementary use of automated and peer feedback has shown potential in existing literature, the integration of these two feedback sources within an interactive learning environment requires further research to be systematically explored. In the current study, the advocacy for an integrative approach to automated

and peer feedback is grounded in the student-feedback interaction model proposed by Lipnevich and Smith (2022), which is reviewed in the following section.

The student-feedback interaction model as the theoretical framework

Lipnevich and Smith (2022) theoretically outlined the mechanism by which feedback enhances learning from a student-feedback interaction perspective. They emphasized that the effectiveness of feedback depends on how feedback is *processed*, rather than merely how it is provided. Their model highlights the need to focus not only on the feedback messages, but also on the agents (i.e., providers and recipients), the context in which the feedback occurs, and the interactions among these components.

The model highlights several key feedback variables, for instance, feedback sources, feedback types, characteristics of the students. Internally, Lipnevich and Smith (2022) incorporated core tenets shared by contemporary feedback researchers. For instance, the route from (multiple) feedback sources to feedback messages reflected their broad definition of feedback as "any information about a performance" that can "come from any source" (p. 2). Their accounts on inclusive feedback sources justified the approaches adopted by previous researchers who attempted to employ automated feedback as a complementary source to human-constructed feedback.

Furthermore, the emphasis on feedback processing in the model aligns with our earlier conceptualization of online dialogic peer feedback. Specifically, the formative and iterative *processes* in online dialogic peer feedback involves student's agentic interactions, use of technological and digital resources to self-generate feedback, collaborative and dialogic sense-making of the feedback information, and active uptake of the feedback for decision-making about next learning steps. Thus, it is logical to position the feedback processing variable in the model as an umbrella construct that supports our latest takes on feedback agency and learner-centeredness, for example,

student-oriented perspective of feedback (Winstone et al., 2022) and the new feedback paradigm (de Kleijn, 2023).

The current study examines the effects of complementing peer feedback with automated feedback on feedback quality and self-regulation. We closely followed the core principles of the student-feedback interaction model to operationalize online dialogic peer feedback, design the feedback intervention (see the intervention section below), and interpret the observed effects.

The current study

Based on the reviewed literature and the theoretical framework, the potential of the complementary use of automated and peer feedback in L2 writing settings is better understood. Nevertheless, an overall evaluation of automated and dialogic peer feedback integration is still needed and anticipated. Therefore, we followed the integration model of AWE-generated and peer feedback proposed by Yao et al. (2021) in our interventional design but shifted the research focus to feedback quality and SRL strategy use among L2 writers. To address the knowledge gap identified in the literature reviewed above, this study is guided by the following research questions (RQ):

(1) To what extent does completing online dialogic peer feedback with automated feedback affect the quality of feedback messages generated by L2 writers?

(2) To what extent does completing online dialogic peer feedback with automated feedback affect L2 writers' use of self-regulatory writing strategies?

Methods and materials

Participants

This study took place in the School of Foreign Languages at a university in China.

Participants were recruited from four classes of a 14-week "Introductory English Writing II" course, taught by the same instructor and following the same curriculum. We applied two recruitment criteria during sampling: (1) participants should have prior experience with and an understanding of formative assessment, particularly dialogic and interactive peer feedback practices; and (2) participants should demonstrate at least an average level of English writing proficiency and sufficient feedback literacy to effectively provide and process peer feedback while regulating their learning behaviors throughout the study. Recruitment decisions were made at the researchers' discretion following a case-by-case review of students' profiles, which included self-reported details relevant to these criteria. Additionally, to preserve the naturally formed peer partnerships, only dyads in which both members qualified and agreed to participate were recruited.

As a result, the researchers recruited a sample of 68 L2 students (M_{age}=18.7 years; 11 males and 57 females). These peer feedback dyads were randomly assigned to either a control group (CG; n=32), using online dialogic peer feedback without automated feedback, or an experimental group (EG; n=36), using the complementary integration of automated and online dialogic peer feedback. Pre-study evaluations indicated that the two groups did not have statistically significant differences in writing proficiency measured by an official TEM-4 writing task [$F(1, 66)=1.56$; $p=.217$; $\eta^2_p=0.02$]. Survey data on prior feedback experiences showed that students commonly engaged in peer feedback (M=3.76 on a five-point scale, SD=1.28); however, only six participants reported prior experience with the integrated use of multiple feedback sources in response to open-ended survey questions.

Intervention

We adopted a theory-informed approach to design the interventional conditions

implemented in this study (see theoretical framework). As illustrated in Figure 1, the two experimental conditions differed only in the feedback sources (i.e., the CG relied solely on peer feedback, whereas the EG utilized both automated and peer feedback) and in how the dyads engaged in dialogic peer feedback discussions to process the feedback. After the dialogic peer feedback sessions, students individually revised their work based on the feedback.

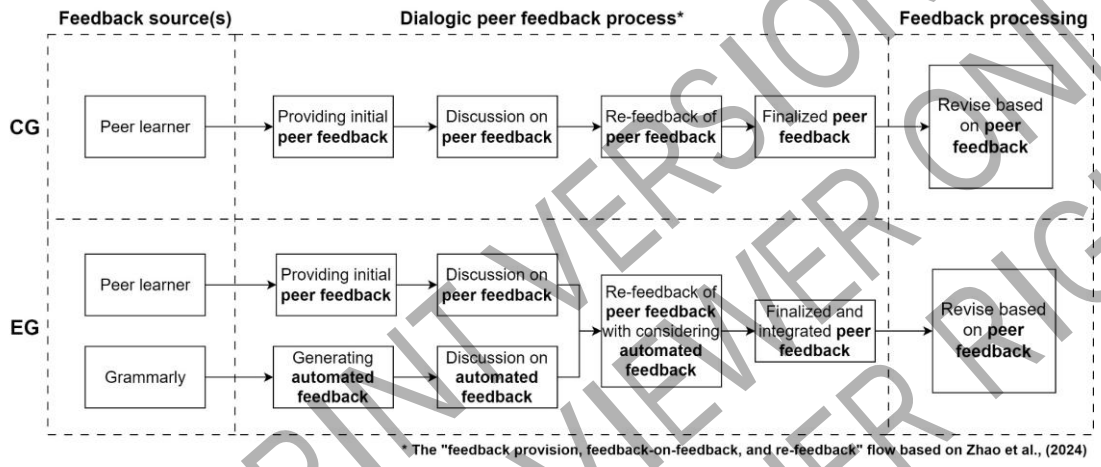

Figure 1. Experimental conditions.

The dialogic peer feedback processes designed for each condition followed the three-step model proposed by Zhao et al. (2024). For the CG, after the initial provision of peer feedback, the feedback dyads were required to engage in a feedback-on-feedback process. The feedback-on-feedback processes involved (1) clarifying the intent of the writing task, (2) seeking clarification on ambiguous feedback points, (3) discussing and resolving disagreements between the feedback provider and recipient, and (4) nailing down alternative revision plans. In the re-feedback stage, the feedback provider refined the original feedback message based on the clarifications and agreements reached during the discussion.

Similarly, EG members engaged in these peer feedback processes but with an additional step: they also discussed and analyzed the automated feedback generated by Grammarly. In the re-feedback stage, EG feedback providers refined their feedback comments by incorporating insights gained from discussions about both peer and automated feedback.

Procedures

This study employed a short-term longitudinal experimental design. As shown in Figure 2, course orientation and dialogic peer feedback training were arranged in the first week. During instructional weeks, the teacher focused on core writing strategies and competence development. In the course, three writing tasks (T1-T3) following the official TEM-4 writing task requirements were assigned. An individual writing and feedback week is allocated for each writing task. Specifically, in each writing and feedback week, students were advised to spend one individual session (45 minutes) to prepare their draft writing, another session to engage in dialogic peer feedback with their partner, and finally one or two sessions to finish the revision based on the received feedback messages.

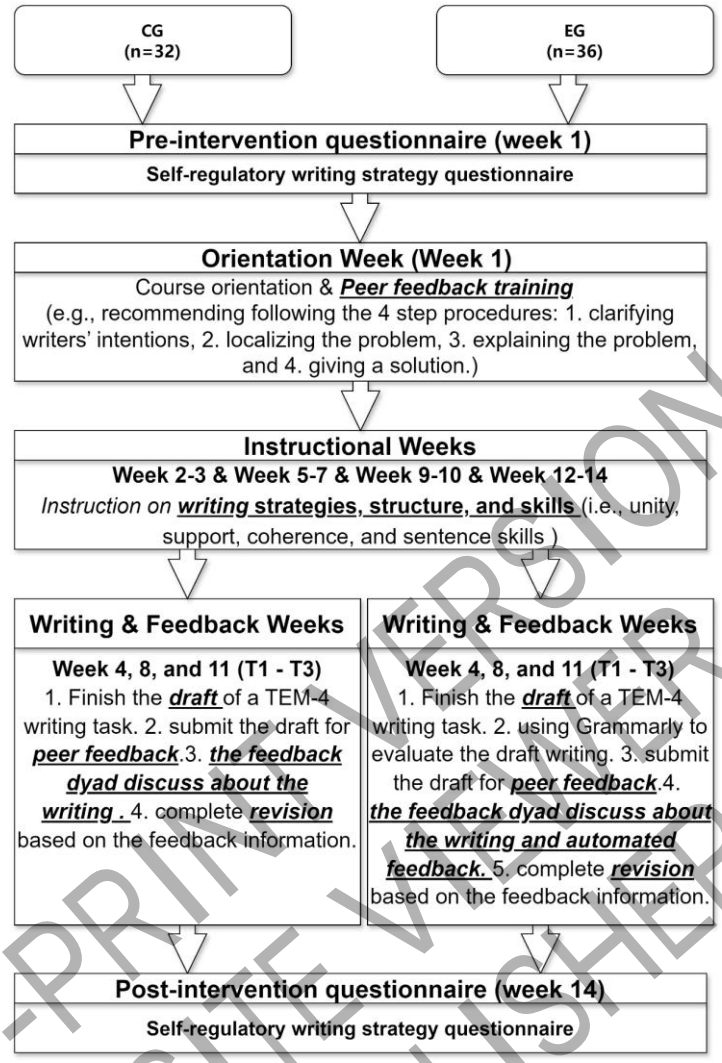

Figure 2. Research procedures.

Dialogic peer feedback was conducted in written forms using Shimo.im, an online collaborative office suite. Feedback-on-feedback and re-feedback were appended to the original feedback messages as discussion threads, enabling a continuous and visible workflow of dialogic and iterative knowledge co-construction. When the dialogic peer feedback processes for a certain writing product concluded, the feedback dyads were requested to submit a finalized version of the feedback messages with which both students were content.

Peer feedback quality was assessed three times, following each writing task. The teacher collected the finalized version of the peer feedback messages for coding and analysis. Self-regulatory writing strategy use was measured through a pre- and post-questionnaire.

Measures

Feedback quality

To measure feedback quality, the coding scheme developed by Kerman et al. (2024) was applied. The coding scheme included three major dimensions of peer feedback quality: (1) affective, (2) cognitive, and (3) constructive. The cognitive category was further divided into three sub-categories: description, identification, and justification. Each of the five features was rated on a scale from 0 to 20, depending on its quality. The details of the coding scheme are shown in **Appendix A**.

To ensure inter-rater reliability, the first author and five co-researchers independently coded 20% of the total peer feedback messages collected from the study. Inter-rater reliability was measured using Fleiss' kappa: (1) affective (0.89), (2) description (0.84), (3) identification (0.79), (4) justification (0.92), and (5) constructive (0.81). Discrepancies were addressed through discussion until consensus was reached. Then, the coders shared the remainder of the coding duty.

Self-regulatory writing strategy questionnaire

We used the SRL writing strategy questionnaire developed by Bai et al. (2022) to assess students' use of self-regulatory writing strategies. The questionnaire included four main dimensions and a total of 17 items. All responses were rated on 5-point Likert-type scale ranging from "never" to "always". The four sections of the instrument respectively

measured students' strategy use planning (4 items), text-generation (4 items), self-monitoring (5 items), and revision (4 items). A sample item under the self-monitoring dimension reads "I re-read the task requirements to generate new ideas." For the research sample, all four sub-scales of the instrument demonstrated high reliability (Cronbach's α = 0.82, 0.92, 0.88, and 0.84, respectively).

Analysis

To address RQ1, Mixed ANOVA tests were conducted using afex[1] and emmeans[2] packages in R software to compare the two conditions in terms of the quality of peer feedback. Before performing the Mixed ANOVA tests, we assessed the assumptions of normality (Shapiro-Wilk test; $p > .05$) and homogeneity of variance (Levene's test; $p > .05$). The Greenhouse-Geisser correction was applied to within-subject factors that violated the sphericity assumption (i.e., $p < .05$ in Mauchly's test). Partial eta squared (η^2_p) was used to assess the effect size in Mixed ANOVA, and values were interpreted based on the suggestions from Gignac and Szodorai (2016).

Following the Mixed ANOVA, post-hoc comparisons were conducted to decompose the interaction effects across different time points. Bonferroni correction was used to adjust *p*-values and control the risk of a Type I error due to multiple hypothesis tests across different time points. Significant pairwise comparisons were followed up with effect size calculations (using Cohen's *d* indices) to assess the magnitude of the observed differences. The interpretation of the Cohen's *d* followed the cutoffs recommended by Plonsky and Oswald (2014) for L2 research .

[1] https://cran.r-project.org/web/packages/afex/index.html
[2] https://cran.r-project.org/web/packages/emmeans/index.html

For RQ2, we performed One-way analysis of covariance (ANCOVA) using afex package to examine whether L2 writers' use of self-regulatory writing strategies improved after the intervention. Prior to conducting ANCOVA, we examined the assumption of homogeneity of regression slopes to ensure that there were no statistically significant interactions between the intervention conditions (treatment; independent variable) and the covariates. The effect sizes for ANCOVA were assessed using partial eta squared (η^2_p) and interpreted following the same standard as in Mixed ANOVA.

Results

Peer feedback quality

Table 1 presents the descriptive statistics of feedback quality, and Table 2 shows the results of the Mixed ANOVA tests. As indicated, statistically significant interactions and main effects (*p*s <.05) of the grouping and test time factors could be found for description, identification, justification, and constructive dimension of feedback quality. However, for the affective dimension, a significant interaction effect was found, but the grouping factor did not have a significant main effect [$F(1,66) = 0.22$; $p = .64$; η^2_p=.003].

Table 1. Descriptive statistics of feedback quality.

Quality	Groups	M (SD)		
		T1	T2	T3
Affective	CG	10.4 (3.2)	11.8 (2.9)	11.7 (3.1)
	EG	10.9 (2.5)	12.1 (3.4)	12.3 (.28)
Description	CG	9.2 (4.1)	10.8 (3.8)	10.1 (2.9)
	EG	11.5 (1.9)	12.8 (4.1)	13.6 (3.2)
Identification	CG	8.9 (2.8)	9.3 (3.1)	10.1 (2.7)
	EG	9.8 (2.1)	12.6 (2.4)	13.4 (2.9)
Justification	CG	7.6 (3.8)	8.1 (2.7)	8.5 (3.6)

Constructive	EG	10.4 (3.1)	11.6 (2.9)	12.7 (3.4)
	CG	7.8 (2.8)	8.6 (3.1)	9.3 (2.6)
	EG	10.4 (4.4)	12.4 (3.6)	14.7 (3.7)

Table 2. Results of Mixed ANOVA tests on peer feedback quality

Quality	Mixed ANOVA				
	Source	df	SS	*F*	η^2_p
Affective	Group	1, 66	4.7	0.22	0.003
	Time	1.47, 97.08	63.6	177.82***	0.729
	Group x Time	1.47, 97.08	3.8	10.76***	0.14
Description	Group	1, 66	322.4	11.87***	0.152
	Time	1.56, 103.23	76.3	48.57***	0.424
	Group x Time	1.56, 103.23	20.9	13.31***	0.168
Identification	Group	1, 66	283.1	16.12***	0.196
	Time	1.88, 124.11	180.6	556.33***	0.894
	Group x Time	1.88, 124.11	50.8	156.48***	0.703
Justification	Group	1, 66	553.4	21.55***	0.246
	Time	1.34, 88.29	78.2	171.15***	0.722
	Group x Time	1.34, 88.29	16.2	35.39***	0.349
Constructive	Group	1, 66	572.6	19.50***	0.228
	Time	1.67, 110.46	338.6	856.83***	0.928
	Group x Time	1.67, 110.46	76.8	194.28***	0.746

Note: *** p<.001; df: Greenhouse-Geisser corrected degrees of freedom; SS: Sum of squares.

Coupled with the descriptive statistics of quality scores of each category across all tests (Table 1), between-subjects post-hoc pairwise comparison showed that: (1) for the affective feature, the CG and EG learners did not have significant differences across the three tasks [ps>.05]; (2) in terms of the description feature, the EG outperformed the CG with large effect sizes in T1 [p<.001; Cohen's d=2.24] and T3 [p<.001; Cohen's d=2.51]; (3) regarding the identification feature, EG surpassed CG in T2 [p<.001; Cohen's d=5.53] and T3 [p<.001; Cohen's d=5.31] with large effect sizes; (4) for justification feature, EG outperformed CG [ps<.001] in all three tests with incrementally larger effect sizes [Cohen's ds=4.02, 4.69, 6.05, respectively]; and (5) pertaining to the constructive feature, EG significantly [ps<.05] surpassed CG in all three tasks by incrementally larger margins [Cohen's ds=2.94, 5.42, 7.76, respectively].

Finally, within-subjects post-hoc pairwise comparisons suggested that both groups made sustained improvement in all five dimensions of feedback quality [*p*s<.05] over time. Specifically, the EG made more visible progress compared to the CG in the identification [Cohen's d_{CG}=2.21; Cohen's d_{EG}=5.86], justification [Cohen's d_{CG}=1.21; Cohen's d_{EG}=3.23], and constructive [Cohen's d_{CG}=2.65; Cohen's d_{EG}=7.48] measures. In contrast, the effect sizes of the overtime progression made by CG [Cohen's d=2.11] and EG [Cohen's d =2.15] were similar for the affective [Cohen's d_{CG}=2.11; Cohen's d_{EG}=2.15] and description [Cohen's d_{CG}=1.01; Cohen's d_{EG}=1.27] features.

Self-regulatory writing strategy use

After controlling for the effects of the pre-questionnaire scores, the results of the ANCOVA test (Table 3) indicated that there were significant differences in scores on text-generation [F(1,65)=6.73; p<.01], self-monitoring [F(1,65)=81.29; p<.001], and revision [F(1,65)=15.81; p<.001] among students in both groups. However, the intervention did not have a significant effect on students' scores in the planning dimension [F(1,65)=0.029; p=.86; η^2_p<0.001]. Moreover, a closer look at the adjusted means and effect sizes shows that the intervention positively exerted small, medium, and large effects on text-generation [η^2_p=0.09], revision [η^2_p=0.19], and self-monitoring [η^2_p=0.56], respectively.

Table 3.ANCOVA results for L2 writers' self-regulatory writing strategy use.

Dimensions	Group	M (SD)		Adjusted Means	SE	ANCOVA	
		Pre	Post			*F*	η^2_p
Planning	CG	3.2 (0.4)	3.4 (0.3)	3.32	0.023	0.029	<0.001
	EG	3.3 (0.6)	3.6 (0.8)	3.32	0.022		
Text-generation	CG	2.8 (0.5)	3.0 (0.4)	2.88	0.026	6.73**	0.09
	EG	2.9 (0.5)	3.3 (0.7)	2.97	0.025		

Self-monitoring	CG	2.9 (0.6)	3.2 (0.8)	2.80	0.063	81.29***	0.56
	EG	2.7 (0.4)	3.8 (1.1)	3.58	0.059		
Revision	CG	3.3 (0.7)	3.6 (0.8)	3.45	0.049	15.81***	0.19
	EG	3.4 (0.6)	4.2 (1.2)	3.72	0.046		

Note: ** $p<.01$; *** $p<.001$

Discussion

In answering **RQ1**, the study demonstrated that the complementary integration of automated and dialogic peer feedback supported students to construct feedback messages with more comprehensive and critical content and constructiveness. The findings were generally in alignment with previous understanding that the inclusion of multiple feedback sources in a technology-mediated language learning setting could improve the quality and usability of feedback (Loncar et al., 2023).

Specifically, students' improvements in content-related feedback quality can be attributed to their enhanced critical thinking abilities (Liu et al., 2023) and increased cognitive engagement in the feedback process (Zhang & Hyland, 2022). Moreover, with the relatively higher level of interactivity during peer feedback processes, the gains in higher-order cognitive skills enable students to provide more constructive comments that facilitate recipients' uptake and utilization (Noroozi et al., 2023). Thus, the findings from this study suggest that integrating automated and dialogic peer feedback can help address persistent criticisms of peer feedback, particularly its tendency toward superficial evaluation and lack of constructive suggestions for improvement (Latifi et al., 2023).

However, the study also found that the intervention did not enhance the affective features of peer feedback. This is somewhat unexpected, as previous literature suggests that technology-mediated dialogs can fortify affective rapport among learners and

"lower relational and social-affective barriers" (Wood, 2022, p. 338). The finding also contrasts with Cheng et al. (2023), who reported that a dialogic peer feedback intervention improved students' trust in peer feedback. In contrast, the present study revealed that, compared to the CG, EG members did not show noticeable improvement in commenting with affectively friendly tones and encouraging emotions.

The discrepancies in research findings could be explained by the methodological differences. While Cheng et al. (2023) measured perceived trust in peer feedback through self-reported questionnaires, the current study analyzed the affective quality of actual feedback comments co-constructed by students in a naturalistic setting. This suggests that establishing "a supportive and secure learning environment" to foster trust and harmony in peer feedback (Lee, 2017, p. 90) may be a complex, time-consuming process that requires the synergy between students' perception and behaviors during feedback processes.

Regarding **RQ2**, this study found that integrating automated and dialogic peer feedback enhanced overall self-regulatory writing strategy use among L2 writers. This finding aligns with previous research on the effects of formative assessment on SRL development (Nicol & Macfarlane-Dick, 2006; Panadero & Lipnevich, 2022).

Notably, EG learners reported higher levels of regulatory strategy use in the during-writing (i.e., text-generation and self-monitoring) and post-writing (i.e., revision) phases. Statistically significant differences between the two groups during writing could be attributed to the belief that dialogic peer feedback can regulate feedback recipients' writing performance and co-regulate feedback providers' evaluative judgment skills (Alemdag & Yildirim, 2022). Similarly, the significant changes in self-reported post-writing SRL strategy use of EG learners could be linked to their expanded opportunities to engage in feedback and feedback-on-feedback sessions, which fostered their self-

reflectivity and subsequent revision decisions (Misiejuk & Wasson, 2021). Overall, this study contributes to existing scholarship by emphasizing that incorporating automated and peer feedback sources significantly enhances self-regulatory writing strategy use.

Unfortunately, the intervention did not significantly impact the use of pre-writing self-regulatory strategies (i.e., planning). This contrasts with the SRL and feedback model (Nicol & Macfarlane-Dick, 2006) and several empirical studies in L2 writing contexts (e.g., Mak & Wong, 2018; Yang et al., 2022, 2023), which suggest that internalizing external feedback can influence the entire SRL process. The lack of statistical significance may be explained by Panadero et al.'s (2016) emphasis on the importance of discussing assessment criteria during the forethought phase of SRL. In this study, however, L2 writers engaging with either peer feedback or automated feedback received a brief training on peer feedback, but assessment criteria were not explicitly provided during the feedback processes. Given this limitation, the intervention could be enhanced by incorporating training or co-development of assessment criteria, as suggested by Bürgermeister et al. (2021) and Author (2024).

Conclusion, limitations and directions for future research

In this study, we investigated the effects of the complementary use of automated and peer feedback on L2 writers' feedback quality and self-regulatory writing strategy use. The findings revealed that the integration of automated and dialogic peer feedback led to improvements in the cognitive and constructive dimensions of feedback quality. Moreover, students exposed to the integrated feedback condition demonstrated enhanced use of self-regulatory writing strategies across key stages of the writing process.

Based on the results of the present study, the following pedagogical implications emerge. First, incorporating computer-generated feedback as a complementary source to

peer feedback should be scaled up in contemporary language classrooms. The study demonstrated that integrating automated and peer feedback improved the quality of student-constructed feedback comments and supported SRL development. However, this integrative approach remains less widely practiced compared to their separate applications. Thus, L2 educators are encouraged to adopt this model to scaffold students' language competence and higher-order skills. Second, educators should pay more attention to students' regulatory abilities. Undoubtedly, the benefits of expanding feedback sources come at the cost of additional time and cognitive demands. To help students effectively engage with these extended learning resources, educators should emphasize skills such as evaluative judgment, feedback literacy, and self- and co-regulation. Finally, educators should design and implement scaffolding measures that enhance students' awareness of and ability to produce, interpret, engage with, and apply feedback, especially from a relational and socio-affective perspective.

Inevitably, this research faced its own limitations. First, the study implemented an integrated approach combining dialogic peer feedback with AWE feedback. Given the emergence of more advanced technology such as generative artificial intelligence systems, AWE systems may no longer be the ideal complement for peer feedback, particularly due to their black-box mechanisms regarding assessment criteria and user customizability. Therefore, future research should explore the possibilities of integrating newer technologies with peer feedback to enhance interactivity in feedback processes. Second, the duration of the intervention and the sample size may still limit its validity and generalizability. Replicating the study with a different sample (in terms of size or origin) and a longer duration or more frequent feedback exposure could yield different and interesting results. Finally, this study employed a quantitative approach to data collection and analysis. Future research could consider incorporating qualitative

approaches such as think-aloud protocols or content analysis to garner nuanced insights about peer feedback.

Additional information

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

<To be added after peer review>

Funding

<To be added after peer review>

References

Alemdag, E., & Yildirim, Z. (2022). Effectiveness of online regulation scaffolds on peer feedback provision and uptake: A mixed methods study. *Computers & Education*, *188*, 104574. https://doi.org/10.1016/j.compedu.2022.104574

Bai, B., Wang, J., & Zhou, H. (2022). An intervention study to improve primary school students' self-regulated strategy use in english writing through e-learning in Hong Kong. *Computer Assisted Language Learning*, *35*(9), 2265–2290. https://doi.org/10.1080/09588221.2020.1871030

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

Cheng, K.-H., Hou, H.-T., & Wu, S.-Y. (2014). Exploring students' emotional responses and participation in an online peer assessment activity: A case study. *Interactive Learning Environments*, *22*(3), 271–287. https://doi.org/10.1080/10494820.2011.649766

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

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edition).

Routledge.
de Kleijn, R. A. M. (2023). Supporting student and teacher feedback literacy: An instructional model for student feedback processes. *Assessment & Evaluation in Higher Education*, *48*(2), 186–200. https://doi.org/10.1080/02602938.2021.1967283
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
Foo, S. Y. (2021). Analysing peer feedback in asynchronous online discussions: A case study. *Education and Information Technologies*, *26*(4), 4553–4572. https://doi.org/10.1007/s10639-021-10477-4
Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069
Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487
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
Kao, C.-W., & Reynolds, B. L. (2024). Timed second language writing performance: Effects of perceived teacher vs perceived automated feedback. *Humanities and Social Sciences Communications*, *11*(1), 1–14. https://doi.org/10.1057/s41599-024-03522-3
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
Lam, R. (2015). Understanding EFL students' development of self-regulated learning in a process-oriented writing course. *TESOL Journal*, *6*(3), 527–553. https://doi.org/10.1002/tesj.179
Latifi, S., & Noroozi, O. (2021). Supporting argumentative essay writing through an online supported peer-review script. *Innovations in Education and Teaching International*. https://www.tandfonline.com/doi/abs/10.1080/14703297.2021.1961097
Latifi, S., Noroozi, O., Hatami, J., & Biemans, H. J. A. (2021). How does online peer feedback improve argumentative essay writing and learning? *Innovations in Education and Teaching International*, *58*(2), 195–206. https://doi.org/10.1080/14703297.2019.1687005
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
Lee, I. (2017). Peer feedback in L2 writing. In I. Lee (Ed.), *Classroom Writing Assessment and Feedback in L2 School Contexts* (pp. 83–103). Springer. https://doi.org/10.1007/978-981-10-3924-9_7
Lenth, R. V. (2017). *emmeans: Estimated marginal means, aka least-squares means* (p. 1.10.7) [Data set]. https://doi.org/10.32614/CRAN.package.emmeans
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

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

Mak, P., & Wong, K. M. (2018). Self-regulation through portfolio assessment in writing classrooms. *ELT Journal*, *72*(1), 49–61. https://doi.org/10.1093/elt/ccx012

Misiejuk, K., & Wasson, B. (2021). Backward evaluation in peer assessment: A scoping review. *Computers & Education*, *175*, 104319. https://doi.org/10.1016/j.compedu.2021.104319

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Noroozi, O., Banihashem, S. K., Biemans, H. J. A., Smits, M., Vervoort, M. T. W., & Verbaan, C.-L. (2023). Design, implementation, and evaluation of an online supported peer feedback module to enhance students' argumentative essay quality. *Education and Information Technologies*, *28*(10), 12757–12784. https://doi.org/10.1007/s10639-023-11683-y

Panadero, E. (2023). Toward a paradigm shift in feedback research: Five further steps influenced by self-regulated learning theory. *Educational Psychologist*, *58*(3), 193–204. https://doi.org/10.1080/00461520.2023.2223642

Panadero, E., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding Self-Regulated Learning Through Self-Assessment and Peer Assessment: Guidelines for Classroom Implementation. In D. Laveault & L. Allal (Eds.), *Assessment for Learning: Meeting the Challenge of Implementation* (pp. 311–326). Springer International Publishing. https://doi.org/10.1007/978-3-319-39211-0_18

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

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

Reynolds, B. L., & Zhang, X. (2023). Medical school students' preferences for and perceptions of teacher written corrective feedback on english as a second language academic writing: An intrinsic case study. *Behavioral Sciences*, *13*(1), 13. https://doi.org/10.3390/bs13010013

Shen, B., & Wang, L. (2024). Development and validation of questionnaire for self-regulated learning writing strategies (QSRLWS) for EFL learners. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2023-0192

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. S. (2012). *afex: Analysis of factorial experiments* (p. 1.4-1) [Data set]. https://doi.org/10.32614/CRAN.package.afex

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

Tan, X., Reynolds, B. L., & Ha, X. V. (2024). Oral corrective feedback on lexical errors: A systematic review. *Applied Linguistics Review*, *15*(3), 1177–1221. https://doi.org/10.1515/applirev-2022-0053

Teng, L. S., & Zhang, L. J. (2020). Empowering learners in the second/foreign language classroom: Can self-regulated learning strategies-based writing instruction make a difference? *Journal of Second Language Writing*, *48*, 100701. https://doi.org/10.1016/j.jslw.2019.100701

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, Y., & Schunn, C. D. (2021). From plans to actions: A process model for why feedback features influence feedback implementation. *Instructional Science*, *49*(3), 365–394. https://doi.org/10.1007/s11251-021-09546-5

Yang, L. (Francoise), Liu, Y., & Xu, Z. (2022). Examining the effects of self-regulated learning-based teacher feedback on english-as-a-foreign-language learners' self-regulated writing strategies and writing performance. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1027266

Yang, L. (Francoise), Zhang, L. J., & Dixon, H. R. (2023). Understanding the impact of teacher feedback on EFL students' use of self-regulated writing strategies. *Journal of Second Language Writing*, *60*, 101015. https://doi.org/10.1016/j.jslw.2023.101015

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

Zhang, J., & Zhang, L. J. (2024). Profiling L2 students' writing self-efficacy and its relationship with their writing strategies for self-regulated learning. *System*, *122*, 103253. https://doi.org/10.1016/j.system.2024.103253

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: Peer feedback quality coding scheme (adapted from Kerman et al., 2024)

Dimensions	Definition	Subdimensions	Rating criteria		
			Good feedback (20 marks)	Average feedback (10 marks)	Poor feedback (0 marks)
Affective	Inclusion of positive emotions such as praise or compliments and negative emotions such as anger or disappointments	-	Included encouraging and positive emotions.	Neutral in emotions.	Included discouraging and negative emotions.
Cognitive	With description, identification, and justification of writing evaluation.	Description	Included a summary statement but only to a large extent.	Included a summary statement but only to a small extent.	None or minimal summary statement.
		Identification	Included explicit and localized identification of problem	Included identification of problem without localization of identified problem	None or minimal explicit identification of problem
		Justification	Included elaborations and justifications of identified problem	Included elaborations but no justifications of identified problem	None or minimal elaborations and justifications of identified problem
Constructive	With recommendations and actionable plans for further improvements	-	Included recommendations and actionable plans.	Included recommendations but no actionable plans.	None or minimal recommendations or actionable plans.