System 136 (2026) 103895

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

A GenAI-supported dialogic model to enhance the functions of peer feedback in EFL writing: a mixed-method interventional study

Da (Alex) Yan [a,b,*], Feng Tian [c], Haitao Li [d], Yu Gao [e], Jinhui Li [e]

[a] School of Foreign Language Studies, Wenzhou Medical University, Wenzhou, China
[b] Wenzhou International Communication Research Center, Wenzhou, China
[c] School of Foreign Languages and Cultures, Ningxia University, Yinchuan, China
[d] International School, Foshan University, Foshan, China
[e] School of Foreign Languages, Xinyang Agriculture and Forestry University, Xinyang, China

ARTICLE INFO

Keywords:
Hybrid intelligent feedback
Dialogic peer feedback
GenAI
Feedback integration
EFL writing

ABSTRACT

This study explores how a GenAI-supported dialogic model of peer feedback (PF) affects EFL writers' writing performance, self-efficacy, and self-regulated learning (SRL). Drawing on recent conceptual insights into dialogic and hybrid intelligent feedback, the proposed model integrated GenAI into PF processes to assist students in generating, discussing, and re-formulating feedback comments. A total of 74 Chinese EFL undergraduates participated in this mixed-methods experimental study, which was embedded within a 17-week writing course. Thirty-six students from a class were assigned to the experimental group (using the GenAI-supported dialogic model), while thirty-eight from another class were assigned to the control group (engaging only in dyadic dialogic PF). The study adopted a three-pronged data collection technique: (1) feedback-informed performance improvements were measured based on draft and revised versions of writing products across three writing tasks; (2) pre- and post-questionnaires were administered to assess changes in PF and writing self-efficacy beliefs; and (3) semi-structured interviews were conducted after the intervention to gather students' reflections on how the new PF model influenced their SRL. Findings revealed that the intervention promoted performance improvements, enhanced PF self-efficacy, and activated SRL at various stages of dialogic PF. However, no significant effects were observed on writing self-efficacy. The evaluation of the model's impact across three key functional areas of feedback highlights the potential of GenAI to systematically enhance PF in an EFL writing context. This study contributes meaningful implications for educators and researchers aiming to incorporate advanced technologies in formative PF.

1. Introduction

Peer feedback (PF), a collaborative and interactive process engaging students in reciprocal evaluation of their peers' work, has been increasingly employed to support English as a foreign language (EFL) writing development (Vuogan & Li, 2023; Zhang & Gao, 2024). However, the success of PF calls for sound domain knowledge, ownership of assessment criteria, and higher-order skills such as

* Corresponding author. School of Foreign Language Studies, Wenzhou Medical University, Wenzhou, China.
E-mail address: alexyan1987@outlook.com (D.(A. Yan).

https://doi.org/10.1016/j.system.2025.103895
Received 20 June 2025; Received in revised form 18 October 2025; Accepted 19 October 2025
Available online 28 October 2025

evaluative judgement, which might not be possessed by all learners (Cheng et al., 2023; García et al., 2024). Therefore, there has been an assortment of scholarly work on scaffolding PF with instructional, social, and technological facilitators (e.g., Bürgermeister et al., 2021; Tseng & Er, 2024; Wood, 2021).

Within the sub-category of technological scaffolding, advances were incrementally introduced to optimize approaches to integrating technologies in PF. Specifically, we started from computer-mediated PF, where technology was employed as a communicative facilitator that activates dialogic interactions (Wood, 2021). Later, automated writing evaluation systems were used to complement PF (Liu et al., 2023; Yao et al., 2021). Recently, generative artificial intelligence (GenAI) has been incorporated to support feedback provision (Guo et al., 2024). A review of the literature reveals an evolving perspective on technology in an assessment landscape—from "a platform" to "a feedback source" and to "non-human agent"—highlighting a growing emphasis on meaningful interactions over mere information availability in the latest discourse of PF. Unfortunately, to date, innovative research incorporating state-of-the-art technologies such as GenAI to systematically influence PF processes and participants remains limited. Additionally, there is an inadequate connection between theory and practice to guide the design and implementation of technology-enhanced PF environments. These gaps limit our ability to theorize about hybrid intelligent feedback systems (Banihashem et al., 2025) and fully understand how technology shapes PF as it rapidly evolves.

To address the above gap, we designed and implemented a GenAI-supported dialogic PF model and examined its effects on cognitive, motivational, and self-regulatory aspects of learning in EFL writing classrooms. The study is innovative not only for leveraging GenAI to strengthen PF effectiveness, but also for its motivation to extend existing theories on learner and feedback interactions (e.g., Lipnevich & Smith, 2022). Moreover, the study responds to the call for more studies on the reception end of PF process (Hornstein et al., 2025), and incorporating GenAI technology into formative feedback loops (Banihashem et al., 2025).

2. Theoretical and empirical backgrounds

2.1. Peer feedback in EFL writing: mapping current research through a tri-functional lens

Given the complexity of language learning and PF, a taxonomical lens is needed to map the current research landscape and specify the issues this study intends to address. Therefore, we draw on the three-dimensional perspectives of feedback functions proposed by Panadero and Lipnevich (2022): learning/performance, motivation/affect, and self-regulated learning (SRL), to guide the review of related work.

There is a consensus that PF positively impacts writing performance, for instance, Vuogan and Li (2023) reported a meta-analytic effect size of 0.73. However, concerns persist about the quality of peer-generated feedback. First, PF providers may fail to offer quality feedback due to limited domain knowledge and feedback literacy (Foo, 2021; Latifi & Noroozi, 2021). Second, recipients might not effectively use PF due to a lack of trust and the necessary evaluative capabilities to act upon peer comments (Ibarra-Sáiz et al., 2020). Notably, a recent meta-analysis (Hornstein et al., 2025) revealed that PF recipients have been less effectively supported compared to PF providers, thereby limiting recipients' abilities to process and apply PF for learning gains. Building on this, questions remain about how PF recipients could shift from (passive) "feedback-as-information" receivers into (positive and agentic) "feedback-as-process" participants to make more performance gains (Winstone et al., 2022).

Research has yielded mixed findings regarding the effects of PF on EFL writers' self-efficacy, a motivational belief that influences learning outcomes and behavioral patterns (Bandura, 2001). On one hand, researchers believe that computer-mediated PF associates with increased writing confidence and intrinsic motivation, reflecting enhanced self-efficacy beliefs (Lee & Evans, 2019; Liu et al., 2023; Tseng & Tsai, 2010; Zhan & Teng, 2025). On the other hand, Ruegg (2018) suggested that PF might be a suboptimal choice for self-efficacy development in comparison with teacher feedback due to a lack of constructiveness. These inconsistencies highlight a debate of interest: does exposure to constructive criticism diminish or enhance self-efficacy? The former view is supported by language educators such as Andrade and Evans (2012) and Hyland and Hyland (2001), while the latter is consistent with socio-constructivist perspectives (e.g., Schunk & Swartz, 1993). Clarifying this issue is particularly relevant in today's PF landscape which is increasingly mediated by socio-technical support. Another gap lies in the limited understanding of learners' self-efficacy in PF. PF self-efficacy pertains to students' confidence in themselves and their peers to effectively deal with feedback, making it a key factor that influences participation and engagement in the reciprocal process (Kasch et al., 2022). Unfortunately, comprehensive investigations into how PF affects both feedback-related and learning-related dimensions of self-efficacy remain scarce.

The SRL function of PF has received less research attention and, to some extent, lacks a strong theory-practice connection. Theoretically, SRL frameworks are central to feedback models, often used to explain how internal/self-feedback is generated and sustainably drives learning (see the reviewed models in Lipnevich & Panadero, 2021). Er et al. (2021) theorized dialogic PF as a collaborative process whose procedures aligned with the transition from socially shared regulation to co-regulation (CoRL) and ultimately to SRL. In contrast, most empirical studies on PF and EFL writers' SRL provided limited insight into the underlying mechanisms or interactional dynamics during SRL-guided PF processes. For instance, Zhan and Teng (2025) found that both computer-mediated and in-person PF promoted longitudinal SRL development in Chinese EFL writing, but showed no significant differences between modalities. Their findings were based on self-reported questionnaires, providing only general trends without unpacking specific SRL dimensions. Yan et al. (2025) provided more detailed results regarding changes in specific SRL dimensions following a technology-enhanced dialogic PF intervention. However, the above studies still fall short of explaining how PF activated students to engage in dialogs, internalize PF, and regulate the revision. Thus, reliance on self-report instruments without deeper exploration of the internal mechanisms of PF's SRL functions has resulted in a misalignment with well-established theoretical frameworks. Given the increasing emphasis on interactivity and collaborative knowledge co-construction in current discourse, this mismatch warrants urgent attention.

In summary, current research in EFL writing highlights the potential of PF to support students' cognitive, motivational, and self-regulatory development. However, further efforts are needed in refining feedback design and conducting empirical investigations in each functional area. These insights guided our evaluation of existing technology-enhanced PF research, the formulation of our research questions and the design of the PF model presented in the following sections.

2.2. Using technology to enhance peer feedback: previous insights

We began by operationalizing technology-*enhanced* PF as the integration of technology-based tools to support and augment the cognitive processes inherent in PF. Consequently, we excluded studies that primarily utilized online platforms solely to mediate and facilitate PF communication (e.g., Wood, 2021; Zhan & Teng, 2025). Our synthesis revealed two main branches of inquiry, which broadly align with recent advancements in technology and the shifting focus of feedback research.

The most common technological approach to supporting PF involves integrating AF as a complement. Zhang and Hyland (2022) constructed an integrated feedback environment that sequentially combined automatic evaluation, PF, and teacher feedback to enhance students' engagement throughout the multi-draft writing process. Their study found that this multi-source feedback integration improved students' engagement with feedback, leading to more meaningful and effective revisions. Likewise, Liu et al. (2023) and Tan et al. (2023) provided more focused insights into technology-enhanced PF by implementing interventions that combined peer assessment with automatic evaluation. Their results showed that technology-enhanced PF improved writing performance, learning motivation, critical thinking, and reduced EFL writing anxiety. In a recent study, Sperber et al. (2025) attempted the combination of AI and peer review, however, the sequential combination of automatic evaluation and PF remained consistent with previous work.

A unique case within the first research strand is the study by Yao et al. (2021), which attempted to embed Pigai-generated AF in PF processes. Unlike earlier studies, the researchers required experimental group participants not only to consider the AF content but also to actively use it as a resource to support meaningful and productive PF discussions. The study found that this interactive integration of AF and PF enhanced language mindsets and motivation. This approach was subsequently extended by Yan et al. (2025), who incorporated Grammarly-generated AF into dialogic PF processes. With additional support for effective dialogic interactions, this more recent short-term longitudinal study reinforced the benefits of integrating AF and PF, particularly in improving PF quality and fostering SRL.

The second research strand explored the use of intelligent technologies as auxiliary learning partners within PF processes. While similar to the approach in Yao et al. (2021), this strand is distinguished by a shift from an anthropocentric to a human-computer interaction perspective on technology-enhanced PF. A notable study in this strand is Guo et al. (2024), who designed an AI chatbot named Eva to assist student in generating PF comment. The experimental results revealed that chatbot support enhanced feedback quality and writing performance. In another study, Yu et al. (2025) investigated a hybrid system that used large language models (LLMs) to categorize and summarize PF in video-based online teacher education. The findings indicated that the system not only enhanced PF quality but also increased reviewers' engagement and reflective depth.

Several lessons are learned from the common themes and evolving trends in the reviewed research. First, effective PF practices in current scholarship emphasize interactivity. Second, advanced technologies such as LLMs and AI chatbots show potential to further enhance PF interactivity by blurring traditional boundaries between human and machinic agency. The former aligns with current conceptualization of feedback, such as dialogic PF (Wood, 2021; Zhao et al., 2024), feedback as process (Winstone et al., 2022), and the student-feedback interaction (Lipnevich & Smith, 2022). The latter, however, foregrounds the pressing need to update theoretical frameworks to keep pace with the ongoing technology-driven changes in PF practices. Consequently, we proposed a model of GenAI-supported dialogic PF, integrating research-informed PF design insights with the latest theoretical frameworks on hybrid intelligence.

2.3. Towards a model of GenAI-supported dialogic peer feedback

We drew on the five principles for designing effective hybrid intelligent feedback systems (HIFS) suggested by Banihashem et al. (2025) to guide the development of our GenAI-supported dialogic PF model.

First, effective HIFS should leverage the complementary strengths of AI and human intelligence. Experimental studies show that integrating technology into dialogic PF, whether as complementary information source (e.g., Yao et al., 2021) or a human-like agent (e.g., Guo et al., 2024), improves feedback quality compared to PF alone. This supports the hypothesis that AI can effectively complement human input in dialogic PF. Second, HIFS is defined by continuous cycles of improvement. This iterative nature mirrors the core process of dialogic PF, which involves students in a three-phase loop of "feedback provision → feedback-on-feedback → re-feedback" (Zhao et al., 2024). Empirical studies following this model (Yan, 2024; Yan et al., 2025; Zhao et al., 2024) demonstrate that iterative PF discussions improve feedback quality, learning performance, and higher-order skills. Unfortunately, existing PF models with AI integration such as Guo et al. (2024) have largely focused on the provision phase, leaving the impact on feedback reception underexplored. Third, HIFS is designed for adaptable and personalized feedback experience. PF has long been valued as a scalable alternative to teacher feedback, offering enhanced personalization with reduced logistical demands (Zhang & Hyland, 2022). Recent studies leveraging GenAI (Guo et al., 2024; Yu et al., 2025) have used tailored prompts or templates to generate context-sensitive and individualized feedback. Finally, well designed HIFS is characterized by shared agency: the active and collaborative commitment shared by human and non-human participants to shape and take ownership of the feedback process. This specific principle aligns with our sustained commitment to support shared agency in PF practices, as evident by the shift from viewing technology as an additional information source to an actively engaged agent in PF processes. Continuous efforts are still needed, however, since the mechanisms by which AI augments and regulates PF participation and learning remain a "black box" that requires further investigation.

Collectively, these theoretical insights and research efforts have established a solid foundation for the GenAI-supported dialogic PF model. This model is underpinned by three core design philosophies rooted in HIFS principles:

(1) Complementary integration: positioning AI as a collaborative partner to support decision-making and content generation in PF.
(2) Balancing personalization and collaboration: allowing space for students to use AI individually when preparing for and reflecting on dialogic discussions.
(3) Human-centeredness: enabling AI augmentation without sacrificing learner agency.

This innovative model is illustrated in Fig. 1. This model closely aligns with the five principles articulated above (Banihashem et al., 2025). Specifically, we positioned GenAI as a supportive partner in dialogic PF who can *complement* learners' productivity without compromising learner proactivity or the integrity of PF processes. This arrangement aims to maximize the balance between human and GenAI *shared agency*. The entire technology-enhanced PF process is characterized by a high level of *iterative* interactivity. We incorporated several material scaffolds to enable *adaptive* and *personalized* GenAI support across different dialogic PF scenarios. The proposed PF model thus outlines the interventional conditions enacted in the present study (see section 4.3).

3. Research questions

Drawing on this proposed model and the insights derived from the literature review, the current study implemented GenAI-supported dialogic PF in an EFL writing classroom to investigate its effects on the three functions of formative PF. Specifically, the following three research questions (RQs) guided this investigation:

RQ1: To what extent does the GenAI-supported dialogic PF affect EFL writers' feedback-informed writing performance improvements?
RQ2: To what extent does the GenAI-supported dialogic PF affect EFL writers' self-efficacy in PF and writing?
RQ3: How does the GenAI-supported dialogic PF affect students' SRL within the PF processes?

4. Method

4.1. Participants

This study was conducted over a 17-week period in an EFL writing course for second-year English majors at a Chinese university. The course, which utilized a unified curriculum and assessment plan and was taught by the same instructor, enrolled 156 students (distributed across four classes: $n_{class1} = 36$, $n_{class2} = 38$, $n_{class3} = 42$, $n_{class4} = 40$). To determine sample size, we conducted an a priori

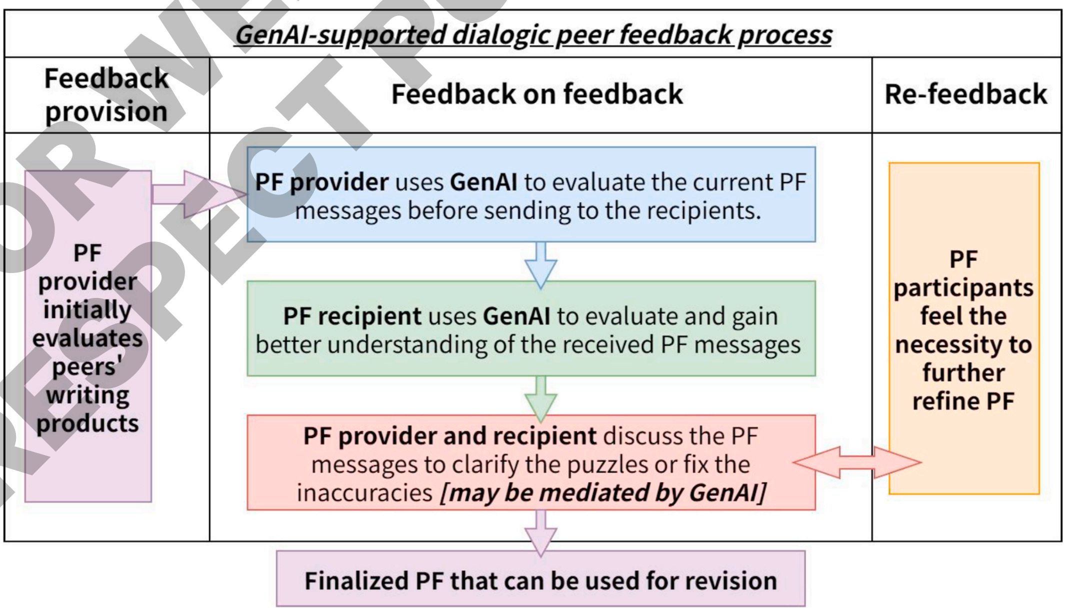

Fig. 1. A GenAI-supported model of dialogic peer feedback.

power analysis using an alpha level of 0.05, a minimum statistical power of 0.80 and considering an effect size of $f = 0.25$ as medium. The analysis indicated a minimal sample of 63 participants in total.

Then we selected students from class 3 ($M_{age} = 19.6$, $n_{female} = 32$) and class 4 ($M_{age} = 16.7$, $n_{female} = 31$) as participants for the following considerations. First, dyadic PF had been used in both the current and prior EFL writing courses taught by the same instructor, and intact sampling helped preserve existing PF partnerships. Second, the two selected classes were relatively large, allowing room for potential dropouts. Third, we conducted a pre-study assessment using an official prompt of Test for English Major band 4 (TEM-4) exam, a widely administered English proficiency test for English majors in China. The results revealed no significant differences between the two classes in writing proficiency [$F(1, 80) = 1.1$, $p = .31$]. Finally, a preliminary survey showed that participants were generally experienced in PF (M = 3.7, SD = 1.3) and technology-enhanced feedback (e.g., Grammarly, Pigai, or ChatGPT) as an individual or alternative source to complement human feedback (M = 3.5, SD = 1.1), though only three dyads reported prior experience combining automated and peer feedback. Two dyads from class 3 withdrew due to personal and medical reasons, and one dyad from each class failed the AI plagiarism check for submitted PF comments, resulting in a final sample of 74 valid participants ($n_{class3} = 36$, $n_{class4} = 38$).

4.2. GenAI policies

The GenAI tool used in this research was Qwen 2.5, a freely accessible LLM developed by Alibaba. As the research site lacked formal policies on GenAI usage, we adopted a liberal, ethical, and human-centered approach, guided by the principles for assessment transformation proposed by the Tertiary Education Quality and Standards Agency (TEQSA, 2023).

As highlighted in the TEQSA document, the growing application of GenAI presents challenges to learning and assessment. For example, excessive or unethical GenAI use in assessment can hinder students' achievement and diminish their engagement. To mitigate these concerns and foster authentic engagement with AI and meaningful learning progression, we implemented the following measures in alignment with the TEQSA guidelines.

First, the critical, ethical, and productive use of GenAI in assessments was mandated. Specifically, all written submissions and PF comments involved in the study were screened for plagiarism of AI-generated content to uphold academic integrity. Second, the PF activities in this study were explicitly designed to support a student-centered learning process. As detailed in the following sections, GenAI played a supportive, rather than substitutive, role in facilitating collaborative peer-based learning.

At various stages of the PF process, both providers and receivers were augmented by GenAI to help develop, understand, and deliver evaluative information. However, the responsibility for making key decisions, such as whether the feedback comment should be further refined or applied for revision, remained with the students.

4.3. Interventions

To examine the effects of the proposed PF model on performance improvements, self-efficacy, and SRL, we designed two interventional conditions: (1) the experimental group (EG) exposed to the GenAI-supported dialogic PF; and (2) the control group (CG) using dialogic PF for formative assessment. Class 3 was assigned to EG while class 4 to CG.

First, both groups engaged in reciprocal peer feedback, ensuring that all students evaluated a peer's work while their own work also peer-evaluated. A tailored rubric was supplied to facilitate initial PF generation (see Appendix A). Second, the two conditions differed only in the dialogic interaction phase. In the EG (see Fig. 1), participants completed three dialogic steps: (1) PF provider and GenAI dialog, (2) PF recipient and GenAI dialog, and (3) a discussion between the PF provider and recipient to co-construct final PF comments. In contrast, the CG only conducted a discussion between provider and recipient. Finally, all students individually revised their writing based on the finalized PF comments.

During the provider-GenAI dialog, the PF provider used GenAI to assess the quality of their initial PF comments. A structured prompting template was developed to guide this process and mitigate potential bias (see Appendix B). The template consisted of four key components: (1) a request for GenAI to assess the PF comments, (2) a rubric of PF quality for reference, (3) the PF comments and (4) the writing products. Notably, the rubric was adapted from Kerman et al. (2024) to provide an evaluative framework for affective, cognitive, and constructive features of PF comments.

For the recipient-GenAI dialog, students were encouraged to discuss iteratively with GenAI to clarify and explore feedback points. While no specific prompting template was provided, students were advised to reference the PF quality rubric when seeking clarification. In particular, the three cognitive subdimension in the rubric—descriptions, identification, and justification (of problems or issues in writing)—facilitated GenAI's transformation of vague or underdeveloped feedback into well-articulated comments.

Finally, to support active engagement and meaningful knowledge co-construction during the provider-recipient dialog, participants received a printed PF regulatory script (Appendix C) adapted from Shao et al. (2024).

4.4. Measures

4.4.1. Feedback-informed performance improvements

Writing tasks were selected from a bank of 25 TEM-4 prompts. The selection procedure followed that of a previously published work (Yan & Gao, 2025), featuring a multi-round Delphi process to screen based on task difficulties, task familiarities, and Flesch readability scores. The accompanied rubric evaluated writing across three dimensions each graded on a 5-point scale: (1) content, (2) organization, and (3) language. This rubric closely aligned with the TEM-4 official writing task structure and competence requirements.

Students completed both draft and revised writing tasks on Shimo.im, a Chinese online collaborative platform. Written dialogic PF was executed via discussion threads on the peer-reviewed draft versions, and revisions were made using tracked changes. Three types of writing artifacts were collected: (1) the clean draft, (2) the peer-reviewed draft with comments and discussion threads, and (3) the revised version. Prior to grading, the revised versions were processed by accepting only those changes directly informed by dialogic PF, while all other revisions were rejected.

The first author and four faculty co-researchers scored the writing products. Both clean drafts and revised versions were blindly rated to ensure anonymity regarding authorship and version status (i.e., draft or revised). Initially, all raters independently scored 50 samples, achieving high inter-rater reliability: content (0.84), organization (0.82), and language (0.93). Given this internal consistency, they then divided the remaining scoring duties. Feedback-informed performance improvement was calculated by subtracting draft scores from revised version scores.

4.4.2. *EFL writers' self-efficacy*

To measure EFL writing self-efficacy, we applied the 7-item "second language writing self-efficacy scale" developed by Rahimi and Fathi (2022). A sample item reads "I present my point of view or arguments accurately and effectively when writing in English". In addition, we incorporated a second dimension of "self-efficacy in PF" by including a 5-item subscale adapted from the "peer-feedback orientation scale" developed by Kasch et al. (2022). A sample item reads "I know how to make the most of peer feedback". Minor wording adjustments were made to the original item descriptions to eliminate the provider-dominant stance. A pilot study with 26 EFL non-participants from the target population suggested satisfactory internal consistency scoes ($\alpha_{writing} = 0.82$ and $\alpha_{pf} = 0.78$ respectively). The instrument is a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

4.4.3. *Self-regulated learning semi-structured interview*

A subset of EG participants were invited for post-intervention semi-structured interviews to explore how the new PF model influenced their SRL patterns and experiences. Guided by the Er et al. (2021) model, the interview questions focused on how individual learners' experience of self- and co-regulation during dialogic PF in this new learning environment. The interview protocol was

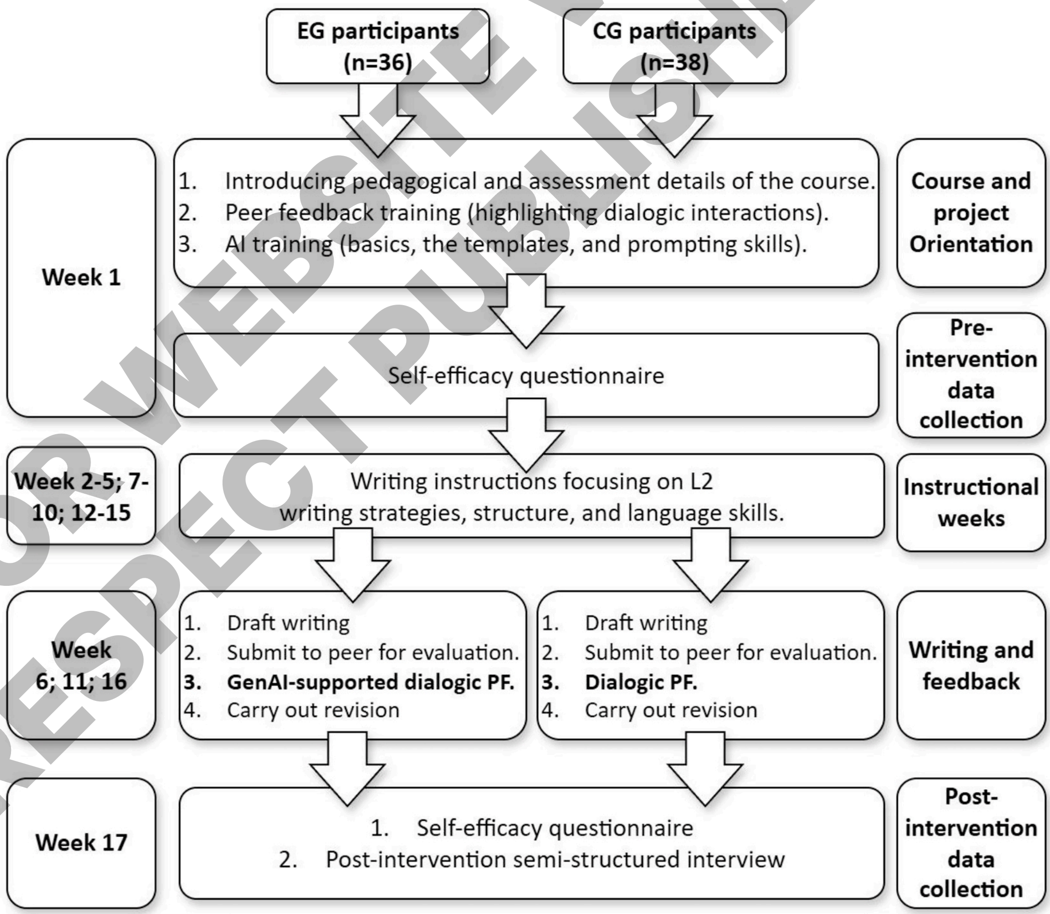

Fig. 2. The experimental procedure.

validated for language quality and trustworthiness by two experts in educational assessment and one industry stakeholder with expertise in human-AI collaboration. Interviews were conducted in Chinese, audio-recorded, and had an average duration of approximately 25 min.

4.5. Procedures

This short-term longitudinal study adopted a mixed-method design (Fig. 2). In Week 1, students attended a brief orientation session to build foundational knowledge about PF and AI use in the course and completed a pre-test questionnaire on self-efficacy. During the instructional weeks, both experimental and control groups, as well as non-participants, received identical writing instruction focused on core writing competencies and language skills.

Over three designated writing and feedback weeks (Weeks 6, 11, and 16), all students completed three writing tasks (T1-T3). Each task spanned several sessions: one 45-min session for draft writing and self-checking, another session for dialogic PF (GenAI-supported PF for the EG vs. only PF for the CG), and one or two additional sessions for revision, reflection, and final edits. A set of sample PF dialogs and comments selected from two dyads belonging to both CG and EG can be found in Appendix D.

In the final week, post-test self-efficacy data were collected, and 12 EG participants were invited for semi-structured interviews. Interviewees were purposefully selected to reflect diverse student perspectives.

4.6. Analysis

Quantitative data were analyzed using R statistical software (version 4.4). For RQ1, a mixed Analysis of Variance (ANOVA) was conducted with writing performance improvement as the dependent variable, test time points (T1-T3) as the within-subjects factor, and intervention group as the between-subjects factor. Prior to analysis, we verified the assumptions of normality (Shapiro-Wilk test, $p > .05$) and homogeneity of variance (Levene's test, $p > .05$). When sphericity was violated (Mauchly's test, $p < .05$), the Greenhouse-Geisser correction was applied. Post hoc comparisons were performed using Bonferroni correction to control for Type I error across multiple tests. In addition, a brief analysis of several PF samples collected from CG and EG dyads was appended to complement and support the statistical results. For RQ2, we performed a one-way analysis of covariance (ANCOVA). The assumption of homogeneity of regression slopes was verified to ensure the absence of statistically significant interactions between the interventional conditions and the covariates. For RQ3, qualitative data from semi-structured interviews were transcribed and analyzed thematically to extract emergent themes from EG participants' narratives regarding SRL changes associated with the new PF model. The data were manually coded and analyzed using an inductive approach, allowing themes to emerge organically without being constrained by a pre-determined theoretical framework.

5. Results

5.1. Effects on performance improvements (RQ1)

Table 1 presents the descriptive statistics of EFL writers' performance improvements after dialogic PF processes for all three tasks. Table 2 shows the result of the mixed ANOVA performed on the performance improvements data. Significant interactional effect between test time (T1-T3) and interventional grouping could be observed for content, language, and total performance improvements, but not for the organization dimension [$F(1.19, 85.94) = 1.42$; $p = .727$]. Regarding the main effects, test time had significant main effects on all four measures. In contrast, interventional grouping showed significant main effects only on organization [$F(1, 72) = 5.34$; $p < .05$] and total [$F(1, 72) = 12.74$; $p < .01$] performance improvements.

Post-hoc analyses on the three wring performance criteria indicated that: (1) EG outperformed CG in organization scores across all three tests [$ts \geq 2.09$; $ps < 0.05$]; (2) there were no statistically significant differences in language measure between the two groups across all three tests [$ts \leq 1.51$; $ps > 0.05$]; (3) EG surpassed CG in content measures only at T3[$t = 3.9$; $p < .01$]; and (4) EG made sustained and significant development [$ts \geq 3.78$; $ps < 0.001$] in all three dimensions from T1 to T3, while CG only made statistically significant improvements from T1 to T3 in content [$t = 2.9$; $p < .05$].

Table 1
Descriptive statistics of performance improvements.

Measure	Groups	M(SD)		
		T1	T2	T3
Content	CG	0.39 (0.64)	0.59 (0.85)	0.49 (0.43)
	EG	0.67 (1.08)	1.06 (1.51)	1.37 (1.29)
Organization	CG	0.39 (0.53)	0.49 (1.06)	0.59 (0.95)
	EG	0.87 (0.86)	1.07 (1.29)	1.26 (1.62)
Language	CG	0.90 (0.21)	1.71 (1.17)	1.03 (0.7)
	EG	0.72 (1.08)	1.15 (1.2)	1.25 (1.41)
Total	CG	1.70 (0.96)	2.80 (0.74)	2.09 (1.28)
	EG	1.88 (0.86)	3.27 (1.08)	3.88 (0.97)

Table 2
Results of the mixed ANOVA on performance improvements.

Variable	Effects	df	F	η_p^2	p
Content	Group	1, 72	5.24	0.068	0.075
	Time	1.60, 115.19	68.76	0.488	<0.001
	Group*Time	1.60, 115.19	37.08	0.340	<0.001
Organization	Group	1, 72	5.34	0.069	.02
	Time	1.19, 85.94	14.08	0.164	<0.001
	Group*Time	1.19, 85.94	1.42	0.019	0.727
Language	Group	1, 72	3.4	0.045	0.69
	Time	1.66, 119.52	23.39	0.245	<0.001
	Group*Time	1.66, 119.52	10.11	0.123	<0.001
Total	Group	1, 72	12.74	0.150	0.002
	Time	1.24, 89.37	869.88	0.924	<0.001
	Group*Time	1.24, 89.37	315.62	0.814	<0.001

Note: df: Greenhouse-Geisser corrected degrees of freedom.

Regarding total performance improvements, post-hoc analysis showed that: (1) regarding total performance improvements, EG outperformed CG at T2 [$t = 2.23$; $p < .01$] and T3 [$t = 6.74$; $p < .001$], with large effect sizes [Cohen's $ds = 1.65$ and 6.16 respectively], but no significant difference was found between the groups at T1 [$t = 0.87$; $p = .39$]; and (2) EG demonstrated a continuous and incremental developmental trajectory across all writing tasks [$ts \geq 9.31$; $ps < 0.001$; $ds \geq 2.08$], whereas CG exhibited an inverted V-shaped reversal in performance improvement, with a substantial gain from T1 to T2 [$t = 31.7$, $d = 3.8$; $p < .001$] followed by a significant decline from T2 to T3 [$t = -11.2$, $d = -2.43$; $p < .001$].

A closer examination of the finalized feedback messages could further illuminate the above statistical results. First, both CG and EG participants adequately identified content issues in peer work, as shown in Appendix D. For instance, CG participants suggested removing redundant arguments (lines 2–5 in A3), while EG participants cautioned against overly subjective arguments ("content" section in B4). However, EG feedback appears more structured and comprehensive, incorporating rationales and evidence-based revision suggestions.

Second, EG feedback provided a more detailed evaluation of organization than CG feedback. A representative example is A3 in Appendix D, where the student commented on both content and language issues in peer work but neglected the aspect of writing organization. Interestingly, we found two entirely different comments on two writing products with highly similar structures (Table 3). Neither PF provider challenged peer performance in organization while drafting their initial PF comments, whereas the EG learner, engaged in GenAI-supported PF, eventually offered critical and actionable revision suggestions for the writer to consider.

Third, PF comments in both groups effectively helped writers identify and rectify issues in language use. As evidenced in A1 and A3 feedback sample (Appendix D), the PF provider in a CG dyad identified two language errors in the first two paragraphs during the initial PF (A1) and supported the writer in correcting them in the finalized PF after discussion (line 6–7 in A3). In contrast, GenAI enabled PF providers to systematically address language use issues when preparing PF comments prior to sharing with peers (see B3 and B4 in Appendix D). As illustrated in the excerpt in Table 4, EG feedback often included a comprehensive list of error corrections across the entire writing artifact. Moreover, EG feedback frequently provided forward-looking suggestions that extended beyond mere error correction.

5.2. Effects on self-efficacy (RQ2)

After controlling for the effects of the pre-test questionnaire scores, the result of the ANCOVA (Table 5) showed a significant and large effect of the intervention on PF self-efficacy scores [$F(1, 71) = 205.6$; $p < .05$; $\eta_p^2 = .74$], but not on writing self-efficacy [$F(1, 71) = 3.72$; $p = .06$]. According to the following-up paired samples t-tests, EG showed a slight decrease [$t = -1.0$; $p = .31$] in writing self-efficacy, while CG showed a slight increase [$t = 0.6$; $p = .55$], both without statistical significance.

Table 3
Excerpts of feedback comments on organization and corresponding feedback interactions.

Excerpt 1: (from a CG dyad)	
PF comment	Corresponding interactions
… the writing is well-structured in a straightforward manner …	Provider praised in initial PF; not discussed/changed in latter phases.
Excerpt 2: (from a EG dyad)	
PF comment	Corresponding interactions
… Though the manner you present your writing is easy to follow, I notice that you tend to keep each part equal in length. This may make the writing less convincible. To improve, you could include more specific details and examples to support your points, as this would make your writing more convincing and impactful …	Provider briefly commented on organization without highlighting any issue in initial PF, AI challenged and elaborated; not discussed; enriched in re-feedback.

Note: Interactions include peer and learner-AI ones throughout the PF process.

Table 4
Excerpted PF comment on language by an EG learner.

Excerpt 3: (from a EG dyad)

Overall language use evaluation

The language use is undermined by recurring grammatical errors, particularly in subject-verb agreement, verb tense consistency, sentence structure, and modifier placement. These issues make some sentences awkward or unclear, but they are common for learners and can be improved with focused practice on grammar rules.

<Other error corrections omitted>

Error 4: Run-On Sentence

- **Identification:** The sentence "Last year, my group was assigned a science project, we finished it quickly because everyone contributed their unique skills" in the second paragraph is a run-on.
- **Explanation:** Two independent clauses are incorrectly joined by a comma without a conjunction, creating a comma splice.
- **How to Correct:** Change to "Last year, my group was assigned a science project, and we finished it quickly because everyone contributed their unique skills."
- **Suggestions:** Study run-on sentences on Purdue OWL and practice splitting or joining long sentences correctly in your writing.

<Other error corrections omitted>

Table 5
ANCOVA results of self-efficacy.

Measures	Groups	M (SD)		Adjusted Means	SE	ANCOVA	
		Pre	Post			F	η_p^2
Writing SE	CG	2.7 (1.1)	2.8 (0.5)	2.92	0.107	3.72	0.05
	EG	2.9 (0.6)	2.7 (1.2)	2.63	0.110		
Feedback SE	CG	2.6 (0.8)	3.4 (1.0)	3.23	0.036	205.6*	0.74
	EG	2.4 (0.7)	3.8 (1.2)	3.98	0.037		

Note: SE: self-efficacy; $^*p < .05$.

5.3. SRL within the GenAI-supported dialogic PF processes (RQ3)

The following sections present the themes identified from the qualitative data. A codebook of the thematic analysis, including alignment with phases of dialogic PF and SRL and supportive salient quotes, can be found in Appendix E. Additionally, during the interviews, we encouraged participants to metaphorically describe the roles played by GenAI in influencing SRL. These “roles” serve as a personification and vivid description of the themes.

5.3.1. GenAI enhances PF generation

A commonly reported benefit of the GenAI-supported dialogic PF model is its enhancement of PF quality. Interview responses indicated that integrating GenAI improved both the cognitive depth (via verification and explanation) and presentation (through gap-filling and expansion) of PF comments. As intended, most improvements occurred during the performance phase of SRL (or the feedback provision and feedback-on-feedback phases in dialogic PF).

GenAI played key roles at multiple stages of the dialogic interactions. After drafting initial feedback, providers used GenAI to verify accuracy (V1) and enrich content (G1) before sharing with recipients. Recipients then engaged in similar processes, that is, verifying understanding (V4) and expanding insights (G2), to build a fuller grasp of the feedback. This often led to requests for further clarification (E3), prompting providers to use GenAI for detailed explanations (E1). The interaction typically concluded with a final PF version, which GenAI could further clarify (E4) and refine (G3). This process highlights GenAI’s potential to enhance the constructive, cognitive aspects of knowledge co-construction in dialogic PF. Crucially, it does not diminish or replace human agency but rather enables learner proactivity and productivity.

Notably, though less frequently observed, GenAI also aided learners during the forethought and reflection phases of SRL. Some used it to prepare for upcoming discussions (V3, E2), while others applied it for self-assessment after revising their work (V5). These uses suggest opportunities to refine the dialogic PF process through better pre-discussion planning and post-feedback reflection.

5.3.2. GenAI substantiates peer dialogs

Another major effect identified in the data is GenAI’s ability to support and stabilize peer dialogues. In addition to mediating interactions between feedback providers and recipients, GenAI also helps ease affective tensions that arise from overly critical or harsh feedback tones.

As shown in code M1, GenAI clarified ambiguities and supported the co-construction of specific feedback points. In human-only settings, such moments might lead to miscommunication or poorly justified feedback. By addressing these issues early, GenAI helped prevent exchanges from becoming “false constructive”, a scenario that could damage PF relationships or discourage future peer engagement. Additionally, GenAI prepared participants both motivationally (M3, M5) and cognitively (M2) for sustained dialog. Notably, unlike the earlier theme where GenAI primarily enhanced individual feedback generation/processing, narratives around dialogic support were largely shared from a collaborative perspective. This suggests a clear, bi-directional influence on both providers and recipients.

Moreover, the GenAI-supported dialogic PF model fostered a more supportive atmosphere for peer collaboration. Students delivered feedback in more acceptable forms (S2) and responded to received feedback with greater criticality and resilience (S1, S4). Together, these emotional and affective benefits strengthen both the process and outcomes of dialogic PF, reinforcing its potential to improve PF quality.

5.3.3. GenAI bridges SRL and CoRL

GenAI has also created new opportunities for PF providers to be co-regulated by both the recipient and AI through insights gained from previous dialogic interactions. Specifically, in Case I1, the provider utilized the PF provision and discussion not only as a means to support a peer's learning but also as an opportunity for her own learning through reflection and self-feedback generation, both enhanced by GenAI support. Similarly, in Case I2, the learner leveraged GenAI's rapid content generation capability to immediately apply feedback prepared for a peer as scaffolding material for evaluating his own writing. Thus, the GenAI-supported model elevated the reciprocal value of PF, transforming it from a peer-based evaluation mechanism into a platform for mutual development.

5.3.4. Negative codes

Interviewees expressed concerns regarding the efficacy of GenAI in supporting successful PF experiences. All negative feedback pertained to the performance phase of SRL. First, participants were dissatisfied with GenAI's tendency to dehumanize PF comments (S3, M4, and V2). While AI-generated comments may seemingly be appealing and content-rich, they might undermine the constructiveness of feedback. Second, several participants raised issues related to content quality, particularly GenAI's propensity to "hallucinate" or provide unreliable comments (V6 and E5). These concerns underscore the vulnerability of technology-rich feedback environments, especially when unrestrained AI usage displaces human agency and critical thinking.

6. Discussion and conclusions

6.1. Effects on performance improvements

The study revealed that both EG and CG participants showed improved performance following the PF across the three writing tasks in the EFL writing course. We further found the EG learners significantly outperformed CG learners in making more effective revisions at T2 and T3. Additionally, EG members sustained significant performance gains from T1 to T3, while the effects of dialogic PF on writing performance tapered off in the latter phases of the study among CG students.

The positive effects of the GenAI-supported dialogic PF model on overall performance improvements can be explained from various perspectives. First, as evidenced in previous research (e.g., Liu et al., 2023), the use of technology in a formative PF setting facilitates greater improvement in EFL writing performance. This study further reinforces the finding that integrating technology into PF is an effective scaffold in contemporary practice. Second, the results are consistent with those of Guo et al. (2024), who implemented an AI-based chatbot to support PF provision in an EFL writing context. In the present study, the systematic integration of AI support in the dialogic PF processes created more opportunities for PF providers and recipients to clarify and elaborate on the feedback. The enhancements in the constructiveness and applicability of the PF comments likely translated into better feedback uptake, as indicated by the improvements in feedback-informed revision demonstrated by the EFL writers (Kerman et al., 2024; Zhao et al., 2024).

Furthermore, the findings about three writing performance criteria (i.e., content, organization, and language) from this study diverge from previous research, such as the study by Guo et al. (2024). Specifically, the current study identified a positive effect of the GenAI-supported dialogic PF model on organization. In contrast, Guo et al. (2024) found that AI-supported PF effectively enhanced content and language performance but had no significant impact on organization. This discrepancy may be attributed to methodological differences between the two studies. Guo et al. (2024)implemented AI support to scaffold PF provision within a non-dialogic PF environment. Moreover, they measured and analyzed the writing performance of PF providers who were exposed to the technology-mediated environment solely during the feedback provision phase. In comparison, EG learners of the current study collaborated in feedback dyads and received GenAI support at both the provider and recipient ends, as well as engaged in peer-based dialogic interactions during the feedback-on-feedback phase. As a result, the findings from the two studies should be viewed as complementary. Shifting the PF recipient from a passive message receiver to an active dialogic participant provided students with greater opportunities to address local issues (e.g., language use) in their writing. Additionally, embedding GenAI support throughout the full dialogic PF process increased recipients' capacity to make more effective, feedback-informed revisions to address global issues (e.g., organization and content) in their writing. Collectively, these findings contribute to a more comprehensive understanding of the effects of AI-supported PF in EFL writing. This study also addresses the recent call for further research into how to better scaffold PF recipients (Hornstein et al., 2025).

Finally, the study also illuminates the longitudinal development of writing performance among cohorts exposed to (dialogic) PF supports. Most previous experimental studies applied PF as a one-off intervention, featuring a pretest-posttest design on writing performance (e.g., Guo et al., 2024; Liu et al., 2023; Zhan & Teng, 2025). Conversely, this study offers valuable insights into the longitudinal trajectories of PF effects. Specifically, the EG demonstrated a stronger propensity for sustained and active uptake of dialogic PF, whereas the CG's trajectory showcased a visible reversal. Thus, we can infer, through the lens of student-feedback interaction (Lipnevich & Smith, 2022), that HIFS increased PF proactivity and receptivity, both of which contributed positively to better feedback engagement. Additional data from the study support this interpretation. For instance, interview responses revealed that EG participants perceived AI as instrumental in sustaining dialog and facilitating more cognitively and motivationally rewarding interactions within PF dyads (see Section 5.3 and the codebook). Moreover, our quantitative analyses of learning data corroborate these observations: EG dyads generated more comment points per session (EG: 9.4 vs. CG: 5.7) and wrote substantially longer PF comments (EG: 121.7 vs. CG: 84.5 words) than their CG counterparts.

6.2. *Effects on self-efficacy*

This study demonstrated that the GenAI-supported dialogic PF model exerted a significantly positive influence on feedback self-efficacy yet failed to produce a positive effect on EFL writing self-efficacy.

The positive effect on feedback self-efficacy observed in this study aligns with findings from previous research (e.g., Cheng et al., 2023). In their study, learners supported by regulatory PF scripts demonstrated increased confidence in both their own and their PF partners' ability to generate high-quality feedback. Exposure to the GenAI-supported dialogic PF facilitated the construction of a socio-affectively supportive environment, which was favorable for learners' emotional wellbeing and fostered trusting relationships between PF participants (Carless, 2012). The above explanation is further supported by the interview insights, that is, students' narratives detailing their enhanced confidence to support their peers' learning and engage in PF dialogues (see 5.3 and the codebook).

A somewhat unexpected finding is that the GenAI-supported dialogic PF model did not exert a significant effect on EFL writing self-efficacy. In prior research, PF has often been identified as an effective strategy for fostering the development of EFL writing self-efficacy (Lee & Evans, 2019; Liu et al., 2023; Tseng & Tsai, 2010; Zhan & Teng, 2025). The discrepancy between the current findings and those of previous studies may be partially attributed to methodological differences. For instance, Lee and Evans (2019), and Zhan and Teng (2025) examined the effects of computer-mediated versus face-to-face PF modes, whereas the present study focused more narrowly on comparing two online dialogic PF conditions (with AI support or otherwise). Following this line of reasoning, it can be inferred that integrating AI into dialogic PF may not yield additional gains in writing self-efficacy, at least within the span of a single semester. This in turn raises another question warranting attention: why didn't dialogic PF promote writing self-efficacy in this study?

At first glance, the Ruegg (2018) study, which found that PF induced no improvement in students' writing self-efficacy, might help to explain the above question. However, since Ruegg (2018) compared the effects of teacher and peer feedback, her explanation based on the quantity of constructive comments may be confounded by factors such as differences in the (perceived) authority of the feedback providers. In contrast, both conditions in this study were peer-based and situated within learning environments that fostered trust and socio-affective support. Consequently, the lack of improvement in writing self-efficacy suggests a different underlying mechanism. A plausible explanation for this stagnation may lie in the overemphasis on "PF constructiveness" in both intervention conditions. Both groups received regulatory scripts to guide their dialogic interactions, while the EG was further scaffolded with a prompt template aimed at enhancing feedback dialog productivity and constructiveness. While these supports appeared effective in boosting feedback self-efficacy, as evidenced by students' increased confidence in engaging in PF dialogues, they may have diminished learners' confidence in their writing abilities due to excessive exposure to critical feedback. According to Andrade and Evans (2012) and Hyland and Hyland (2001), extensive access to constructive feedback comments may decrease EFL writers' confidence in writing. As a reflective consideration, it is worthwhile to revisit the "how much is too much" question raised by Ruegg (2018) by seeking a balance between fostering critical engagement and supporting learners' psychological well-being.

6.3. *Effects on self-regulated learning*

This study identified various supportive effects of the GenAI-supported model on SRL of EFL writers engaged in dialogic PF. Overall, this new model of PF is likely to become more productive, interactive, and reciprocal, compared to the human-only settings.

First, our data indicated that GenAI supports meaningful dialogic PF generation. It became evident that GenAI support was available across all three key phases of a dialogic PF loop (Zhao et al., 2024), indicating that it plays a significant role in enhancing the traditionally student-led process of providing and responding to feedback. This GenAI-supported approach offers a promising response to longstanding concerns that the effectiveness of peer feedback is often limited by students' insufficient domain knowledge and underdeveloped feedback literacy (Foo, 2021; Latifi & Noroozi, 2021). These positive effects can be attributed to the expanded participatory scope and interactive patterns within PF collaboration, which enable shared agency and AI augmentation across diverse dialogic contexts. Drawing on Zimmerman's (1989) triadic model of SRL, we observe that changes in the learning environment, such as the integration of AI in PF, trigger a reconfiguration of individual SRL processes. For instance, students increasingly consult GenAI before or after engaging in peer-based feedback discussions, a shift that reflects deeper transformations in their self-regulatory strategies. These include more deliberate goal setting, strategic planning, and reflective thinking, all of which are essential components of effective SRL in AI-enhanced collaborative settings.

Second, this study indicated that GenAI support helped maintain students' stable mindset, willingness, and confidence in engaging with dialogic PF. This phenomenon can be interpreted through Boekaerts' (2011) dual-processing model of SRL, which emphasizes that learners' emotional responses and choice of goal pathways are shaped by their appraisal of the learning context. In this study, the personalized scaffolding provided by GenAI helped align the PF process with students' individual learning goals and emotional needs, thereby reducing anxiety and fostering a sense of control and competence. As a result, students were more emotionally equipped to actively participate in dialogic PF, which in turn supported PF self-efficacy growth and feedback-informed performance improvements. The interplay between these cognitive and emotional dimensions aligns with Carless' (2012) assertion that a supportive emotional climate enhances learners' trust and confidence, which ultimately lead to more meaningful engagement with feedback.

Finally, interviewees reported extended CoRL that can support SRL. In current scholarly discourse, a notable advantage of dialogic PF lies in its reciprocity, that is, the occurrence of CoRL during dialogic interactions. Traditionally, scholarly attention has predominantly focused on provider-initiated CoRL within dialogic PF (e.g., Er et al., 2021). In contrast, Zhu and To (2022) offered seminal insights into the other side of this mechanism, that is, how PF receivers co-regulate providers' learning through back-evaluation and resultant discussion. This study added new insights into how PF providers are co-regulated in a technology-enhanced dialogic setting. Specifically, in addition to being co-regulated by PF recipient during dialogic interactions, the providers might also be co-regulated by AI during post-PF reflection through AI-supported PF comment rechecking, or direct application of the feedback as evaluative criteria of self-assessment. In light of Hadwin et al.'s (2018) viewpoints that regulatory expertise is distributed across individuals in collaborative learning, the GenAI-supported dialogic PF model has the potential to foster a mutually beneficial learning environment by expanding the patterns of co-regulatory scaffolding.

6.4. Pedagogical implications

The study holds several implications for higher education educators interested in technology-enhanced PF. First, it demonstrates that GenAI-supported dialogic PF has significant potential to enhance multiple aspects of language learning outcomes. Reflecting on our previous scholarly work in this field highlights a promising path forward: while dialogic PF contributes meaningfully to interactivity and constructiveness in feedback processes, it may not be sufficient on its own to support sustained learning advancement. Therefore, we recommend integrating instructional scaffolding with dialogic PF to achieve more comprehensive and enduring outcomes in higher education. Noticeably, the implication should not be confined to language learning, as our approach of GenAI-supported dialogic PF could be adapted for other disciplinary areas easily.

Second, with emergent technological advancements, it is opportune to further investigate HIFS, which can foster deeper learning through more interactive and collaborative modes of engagement. The findings suggest that HIFS can support meaningful learning progression without causing passive or defensive engagement with feedback. Unfortunately, such innovations remain under-researched and underutilized in both language learning and higher education more broadly. Therefore, we recommend adopting a human-centered and interactive approach to GenAI in feedback and assessment practices within higher education, aiming to strike a balance between AI augmentation and learner proactivity.

Finally, the study revealed that researchers and educators in educational assessment should devote more attention to feedback designs. We attribute the outcome of this study to our feedback design philosophy: embedding technological enhancement within a student-centered and agency-oriented framework of formative assessment. In recent months, we have observed a growing surge in research interest in technology-enhanced feedback. However, despite this increased attention, feedback designs that are both elegant and grounded in robust theoretical frameworks remain relatively scarce. Therefore, we recommend that scholars expand existing theories by integrating insights derived from empirical research.

6.5. Limitations and future directions

The study is among the earliest inquiries into GenAI-supported dialogic PF with a higher education and language learning context. However, it faces several limitations that would guide future research. First, we adopted a tri-functional perspective to identify the dependent variables involved in this experimental study (Panadero & Lipnevich, 2022). This choice provided a relatively comprehensive lens through which to evaluate the impact of the new PF model on learning. Nevertheless, this approach may have limited the depth of investigation within each functional dimension. For instance, regarding the motivational/affective aspect of learning, the study revealed mixed effects across different self-efficacy measures. Unfortunately, further investigation into the reasons underlying these findings would have rendered the current paper excessively overwhelming for its readers. Therefore, further exploration of the nuanced mediating mechanisms of each function of technology-enhanced PF is warranted.

Second, the current interventional study focused primarily on externally observable learning outcomes rather than internally observable ones (Nicol & Macfarlane-Dick, 2006). For example, the paper does not examine how the innovation affected the content and quality of PF comments before and after GenAI enhancement. Similarly, the study did not explore the relationship between individual characteristics and feedback dynamics across different PF dyads. These limitations constrain our comprehensive understanding of the nature of feedback within PF processes, as well as the influence of individual and contextual factors.

Third, in the current study, we provided learners in the experimental group with various material scaffolds to support their use of GenAI and engagement with PF. Upon analysis and reflection, we realized that the current configuration has room for improvement. Specifically, there appeared to be an overemphasis on feedback constructiveness, which, while beneficial for learning outcomes, might pose risks to learners' long-term psychological well-being. In the future, researchers might delve deeper into the feedback des specifications, for example, how to finetune the allocation of different scaffolding resources in a hybrid intelligent learning context.

Finally, the study might be limited by its relatively small sample and the lack of sample heterogeneity. The sampling strategy is understandable given our primary objective to achieve a holistic evaluation of the new PF model. Recruiting students from a learning environment that is easily accessible and familiar to the researchers was therefore a reasonable decision. Future research should aim to expand both the size and diversity of the sample to enhance the generalizability of the findings.

CRediT authorship contribution statement

Da (Alex) Yan: Writing – review & editing, Writing – original draft, Data curation, Conceptualization. **Feng Tian:** Writing – original draft, Conceptualization. **Haitao Li:** Writing – original draft. **Yu Gao:** Writing – original draft. **Jinhui Li:** Writing – original draft.

Statement of GenAI usage

We used the Qwen3-235B-A22B large language model to proofread this paper. Artificial intelligence was not used to generate any new content, and all suggested amendments were first reviewed by the first author before being applied.

Conflict of interest

The authors have no conflicts of interest to declare.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.system.2025.103895.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Andrade, M. S., & Evans, N. W. (2012). *Principles and practices for response in second language writing*. Routledge. https://doi.org/10.4324/9780203804605

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*, 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Banihashem, S. K., Noroozi, O., Khosravi, H., Schunn, C. D., & Drachsler, H. (2025). Pedagogical framework for hybrid intelligent feedback. *Innovations in Education & Teaching International*. https://doi.org/10.1080/14703297.2025.2499174

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 408–425). Routledge.

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

Carless, D. (2012). Trust and its role in facilitating dialogic feedback. In *Feedback in higher and professional education* (pp. 90–103). Routledge.

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

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

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*, 1–14. https://doi.org/10.1080/02602938.2024.2329170

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*, Article 100962. https://doi.org/10.1016/j.iheduc.2024.100962

Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In *Handbook of self-regulation of learning and performance* (2nd ed., pp. 83–106). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-6.

Hornstein, J., Keller, M. V., Greisel, M., Dresel, M., & Kollar, I. (2025). Enhancing the peer-feedback process through instructional support: A meta-analysis. *Educational Psychology Review, 37*(2), 42. https://doi.org/10.1007/s10648-025-10017-3

Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing, 10*(3), 185–212. https://doi.org/10.1016/S1060-3743(01)00038-8

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

Kasch, J., Van Rosmalen, P., Henderikx, M., & Kalz, M. (2022). The factor structure of the peer-feedback orientation scale (PFOS): Toward a measure for assessing students' peer-feedback dispositions. *Assessment & Evaluation in Higher Education, 47*(1), 15–28. https://doi.org/10.1080/02602938.2021.1893650

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. (2021). Supporting argumentative essay writing through an online supported peer-review script. *Innovations in Education & Teaching International, 58*(5), 501–511. https://doi.org/10.1080/14703297.2021.1961097

Lee, M., & Evans, M. (2019). Investigating the operating mechanisms of the sources of L2 writing self-efficacy at the stages of giving and receiving peer feedback. *The Modern Language Journal, 103*(4), 831–847. https://doi.org/10.1111/modl.12598

Lipnevich, A. A., & Panadero, E. (2021). A review of feedback models and theories: Descriptions, definitions, and conclusions. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.720195

Lipnevich, A. A., & Smith, J. K. (2022). Student – Feedback interaction model: Revised. *Studies In Educational Evaluation, 75*, Article 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

Nicol, D., & 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

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

Rahimi, M., & Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: A mixed methods study. *Computer Assisted Language Learning, 35*(9), 2627–2674. https://doi.org/10.1080/09588221.2021.1888753

Ruegg, R. (2018). The effect of peer and teacher feedback on changes in EFL students' writing self-efficacy. *Language Learning Journal, 46*(2), 87–102. https://doi.org/10.1080/09571736.2014.958190

Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18*(3), 337–354. https://doi.org/10.1006/ceps.1993.1024

Shao, J., Cheng, L., Wang, Y., Li, K., & Li, Y. (2024). 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

Sperber, L., MacArthur, M., Minnillo, S., Stillman, N., & Whithaus, C. (2025). Peer and AI review + reflection (PAIRR): A human-centered approach to formative assessment. *Computers and Composition, 76*, Article 102921. https://doi.org/10.1016/j.compcom.2025.102921

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

TEQSA. (2023). Assessment reform for the age of artificial intelligence. https://www.teqsa.gov.au/guides-resources/resources/corporate-publications/assessment-reform-age-artificial-intelligence.

Tseng, S.-S., & Er, E. (2024). Maximizing the impact of dialogic peer feedback on student engagement: The role of regulation support. *Educational Technology & Society, 27*(2), 133–148.

Tseng, S.-C., & Tsai, C.-C. (2010). Taiwan college students' self-efficacy and motivation of learning in online peer assessment environments. *The Internet and Higher Education, 13*(3), 164–169. https://doi.org/10.1016/j.iheduc.2010.01.001

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. (2021). A dialogic technology-mediated model of feedback uptake and literacy. *Assessment & Evaluation in Higher Education, 46*(8), 1173–1190. https://doi.org/10.1080/02602938.2020.1852174

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/10.1080/02602938.2024.23330053.

Yan, D. A., & Gao, Y.. Effects of interactive vs. Complementary approaches to integrating automated and peer feedback on L2 writers' feedback quality and performance improvement. Computer Assisted Language Learning. https://www.tandfonline.com/doi/abs/10.1080/09588221.2025.2569352.

Yan, D. A., 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/10.1080/10494820.2025.24916202.

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*, Article 100568. https://doi.org/10.1016/j.asw.2021.100568

Yu, J., Yu, S., & Chen, L. (2025). Using hybrid intelligence to enhance peer feedback for promoting teacher reflection in video-based online learning. *British Journal of Educational Technology, 56*(2), 569–594. https://doi.org/10.1111/bjet.13559

Zhan, Y., & Teng, M. F. (2025). Assessing English writing self-efficacy beliefs, self-regulation, and performance through asynchronous computer-mediated feedback and face-to-face peer feedback. *Educational Assessment, 30*(1), 4–20. https://doi.org/10.1080/10627197.2025.2452442

Zhang, Y., & Gao, Y. (2024). Exploring the dynamics of student engagement with receiving peer feedback in L2 writing. *Assessing Writing, 60*, Article 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*, Article 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

Zhu, Q., & To, J. (2022). Proactive receiver roles in peer feedback dialogue: Facilitating receivers' self-regulation and co-regulating providers' learning. *Assessment & Evaluation in Higher Education, 47*(8), 1200–1212. https://doi.org/10.1080/02602938.2021.2017403

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329