Learning and Instruction 107 (2027) 102489

Contents lists available at ScienceDirect

Learning and Instruction

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

From revision to learning transfer: Self- and socially shared-regulated pathways in AI-assisted feedback

Shijun (Cindy) Chen [a,*], Conrad Borchers [b], Christian D. Schunn [c]

[a] Faculty of Education, The University of Hong Kong, Hong Kong, China
[b] Vanderbilt University, Nashville, TN, USA
[c] Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, USA

A B S T R A C T

Background: Generative artificial intelligence (GenAI) has transformed feedback provision in writing practices. However, little is known about the mechanisms through which AI-mediated engagement and revision contribute to longer-term learning transfer.
Research aims: Framed by theories of self-regulated learning (SRL), socially shared regulated learning (SSRL), and cognitive engagement, the study examined the relationship among engagement, immediate revision gains, and longer-term learning transfer in AI-assisted feedback.
Methods: This 16-week cluster-randomized experiment involving 108 undergraduates in China compared two feedback modes: AI-assisted self-feedback and AI-assisted peer feedback. Analyses drew on 1100 writing samples (including pretests, four assignments with two drafts each, and posttests), and 432 reflection forms on AI feedback.
Findings: Linear mixed models revealed a decoupling of learning and performance: while both conditions produced similarly strong second drafts, AI-assisted peer feedback demonstrated significantly greater learning transfer to subsequent tasks. In AI-assisted self-feedback, learning followed a performance-dependent SRL cycle in which AI engagement and revision improvements independently predicted learning. A threshold-like pattern further showed that benefits emerged only when students engaged constructively, measured by reflection forms, to internalize writing standards. Conversely, AI-assisted peer feedback was associated with stronger learning transfer that was less dependent on revision gains or engagement with AI, a pattern consistent with an SSRL perspective emphasizing interactive engagement and co-constructed understanding.
Conclusion: These findings highlight the importance of revision mechanisms and depth of engagement across feedback contexts. The study contributes to theories of regulation and cognitive engagement and offers pedagogical and design implications for fostering meaningful learning through AI-assisted feedback.

1. Introduction

Feedback is essential to good writing (Graham, 2018), as it supports revision and contributes to overall improvement across multiple drafts (Cui et al., 2022; Paulus, 1999). However, its effectiveness depends on multiple variables, including feedback design (Lira Lorca et al., 2026; Peltzer et al., 2024), various feedback sources and writing contexts (Hattie & Timperley, 2007; Scherer et al., 2024), and the extent to which students engage with feedback through ongoing dialogue and reflection. (Winstone & Carless, 2020; Winstone et al., 2017). Recent research on generative artificial intelligence (GenAI) feedback has begun to move from focusing on its technical capacity for providing enriched, personalized, and immediate feedback (e.g., Steiss et al., 2024) toward understanding how students engage with and use such feedback to support learning (Corbin et al., 2025; Winstone et al., 2026). Such engagement can initiate deeper processing of feedback and facilitate the transfer of learning to subsequent tasks (Carless & Boud, 2018).

Empirical studies have shown that, with effective support of GenAI in feedback processes, individual students can engage cognitively and metacognitively by evaluating or extracting AI-generated information, thereby supporting self-regulated learning processes (Chen & Carless, 2026; Chen & To, 2026; Tam, 2025; Zhan et al., 2025; Zhan and Yan, 2026). In peer feedback contexts, GenAI has also been found to enhance students' writing performance and feedback-related skills (e.g., Lu et al., 2026; Noroozi et al., 2025; Yan et al., 2026). Despite these advances, comparative evidence across GenAI-assisted feedback contexts remains limited, particularly regarding how different designs shape students' engagement, revision practices, and learning transfer. This is especially important because GenAI can provide highly capable forms of feedback, which novice students may over-rely on during learning (Fan et al., 2025; Jacobs et al., 2025). Moreover, while earlier studies have mainly focused on immediate outcomes within a specific context, the mechanisms by which engagement or revision processes contribute to learning remain underexplored. Addressing these gaps is

* Corresponding author. Faculty of Education, The University of Hong Kong, Hong Kong, China.
E-mail addresses: shijunchen@connect.hku.hk (S.(C. Chen), c.borchers@vanderbilt.edu (C. Borchers), schunn@pitt.edu (C.D. Schunn).

https://doi.org/10.1016/j.learninstruc.2026.102489
Received 28 May 2026; Received in revised form 15 September 2026; Accepted 5 October 2026
Available online 10 October 2026
0959-4752/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

important because they help clarify what contributes to longer-term learning rather than examining immediate task performance improvement, and they inform effective design of AI-assisted feedback.

This study aims to investigate two modes of AI-assisted feedback: AI-assisted self-feedback (AISF) and AI-assisted peer feedback (AIPF), from a regulatory learning perspective. AISF is framed within a self-regulated learning (SRL) cycle, whereas AIPF is conceptualized as a form of socially shared regulation (SSRL). Although engagement and revision are known to benefit writing improvement, successful revision does not necessarily translate into longer-term learning transfer, as reflected in performance on subsequent writing tasks (Peltzer et al., 2024). This suggests that immediate revision performance alone may not fully capture underlying learning mechanisms and that the quality of students' engagement with feedback plays a critical role. To better account for this, we adopt a more fine-grained perspective on engagement by drawing on the ICAP framework (Chi & Wylie, 2014).

In sum, this study examines how AISF and AIPF differentially shape students' engagement, immediate revision performance, and longer-term learning transfer, and investigates how engagement and revision processes contribute to learning across these contexts.

2. Literature review

2.1. Empirical studies of GenAI in self and peer feedback contexts

In writing contexts, GenAI can generate immediate, scalable, and personalized feedback information (e.g., Debets et al., 2025; Steiss et al., 2024). However, it is more than a tool for feedback delivery; rather, it creates a landscape that can foster a dialogic, student-centered, or process-oriented understanding of feedback (Corbin et al., 2025; Winstone et al., 2026). With strategic implementations of GenAI, student agency can be enhanced within iterative feedback cycles (Chen & Carless, 2026; Darvishi et al., 2024). Here, we make a strong contrast between GenAI-supported feedback, in which AI supplements and integrates with other feedback sources, and direct GenAI feedback, in which students simply act on received GenAI feedback without integrating with other sources, which can lead to limited learning or de-skilling (Winstone et al., 2026). Accordingly, GenAI is increasingly embedded in process-oriented feedback contexts, such as self (or internal feedback) and peer feedback, to support more proactive learner engagement and feedback use (Noroozi et al., 2025; Zhan et al., 2025). Prior research has shown that feedback design and contexts can influence how students engage with and use feedback (e.g., Winstone et al., 2017; Wood, 2022), which in turn translates into performance, suggesting that different AI-assisted feedback approaches may lead to distinct patterns of engagement, revision, and learning.

In peer feedback contexts, Yan et al. (2026) examined a GenAI-supported dialogic model of peer feedback. They found significant improvements with effect sizes Cohen's $ds = 1.7$ and 6.2 for the second and third task respectively in writing performance compared with the control group, alongside enhanced self-regulated learning strategies, including goal setting, strategic planning, and reflection. Other studies further show that GenAI can support the provision of peer feedback by enhancing peer feedback quality with effect size $r = 0.608$ in a Wilcoxon signed-rank test (Guo et al., 2025), and strengthening students' feedback literacy, particularly their ability to act upon feedback (e.g., Lu et al., 2026, with a Partial $\eta^2 = 0.09$). In contrast, empirical evidence on GenAI in self-feedback contexts remains limited. Existing qualitative findings nevertheless suggest that GenAI can facilitate internal feedback processes, such as judgement, reflection, and self-assessment, while also supporting SRL strategies (Chen & To, 2026; Tam, 2025). For example, Chen and Carless (2026) found that a GenAI-supported internal feedback approach encouraged students to engage more proactively with GenAI during revisions, which can strengthen self-reflection across tasks.

The relative benefits of AI-integrated feedback likely depend upon how students engage with AI feedback. Zhan and Yan (2026) identified multiple forms of engagement when students interact individually with ChatGPT, including cognitive engagement (e.g., selective attention, extracting key information, and comparison) and metacognitive engagement (e.g., monitoring, goal setting, and reflection as key SRL strategies). Collectively, these studies suggest that when GenAI is embedded within structured learning processes (e.g., draft-AI use-revision cycles) and student-centered feedback approaches (self-feedback and peer feedback), improvements in writing performance, proactive feedback engagement, or SRL processes are likely to occur, especially when students demonstrate higher levels of engagement with the AI-feedback.

However, self-feedback and peer feedback, even without GenAI assistance, are underpinned by fundamentally different self-regulated learning structures, with peer feedback involving shared-regulation and self-feedback relying more heavily on internally generated regulation (Allal, 2020; Hadwin et al., 2018). Therefore, to understand the similarities and differences in the benefits of GenAI across feedback contexts, an integrative theoretical framework is needed to explain how students engage with GenAI feedback and how they use it to revise and learn across feedback contexts.

2.2. Theoretical underpinnings on GenAI-assisted self-feedback and GenAI-assisted peer feedback

Self-regulated learning theory conceptualizes learning as a cyclical process in which learners actively set goals, monitor performance, and reflect on outcomes to guide subsequent action (Zimmerman, 2000), which comprises three interrelated phases: forethought (e.g., task analysis and motivational belief), performance (e.g., self-control and self-monitoring), and reflection (e.g., self-reflection). Self-feedback, framed within a self-regulated learning perspective, is described as a continuous comparative process in which students use various external information to generate internal feedback and to close the gap between current and desired performance (Butler & Winne, 1995; Lipnevich & Smith, 2022). We then define GenAI-assisted self-feedback (AISF) as a cyclical process in which students set learning goals, engage with GenAI feedback by evaluating it in relation to their goals, needs, and current performance, and reflect on this process to inform subsequent revisions.

Scholars have gone beyond self-regulation and expanded it to the notion of socially shared regulation (SSRL). It refers to how group members jointly regulate their collective activity by constructing shared goals and task perceptions (Panadero & Järvelä, 2015). For example, peer feedback, when implemented dialogically, is a collaborative meaning-making process in which students and their peers discuss feedback to support its uptake and translate it into task engagement (Er et al., 2021). We therefore define GenAI-assisted peer feedback (AIPF) as a collaborative process in which students jointly interpret and evaluate GenAI feedback through peer discussion, monitor their engagement with GenAI feedback, and make collective decisions about its uptake during the revision process.

Building on these conceptualizations, engagement, revision, and learning are treated as interrelated processes that capture how SRL and SSRL regulation phases (i.e., forethought, performance, reflection) unfold in AI-assisted feedback contexts. We examine three key phases of feedback use across tasks: 1) engagement reflects how students interpret, reflect (Handley et al., 2011), and evaluate feedback (Tai et al., 2018; To, 2022); 2) revision captures how they act on feedback to improve immediate revision performance; and 3) learning is reflected in longer-term learning transfer, operationalized as performance beyond

Process	AISF (SRL)	AIPF (SSRL)	Function
Engagement (Forethought & Reflection)	Students *individually* interpret and evaluate GenAI feedback	Students *discuss* and *jointly* interpret and evaluate GenAI feedback	How students understand feedback
Revision (Performance)	Students decide how to revise *on their own*	Students decide how to revise *through peer discussion*	How students use feedback
Learning (Reflection-Next forethought)	Students improve *independently*	Students improve based on *both individual and peer-supported understanding*	How students improve over time (next draft)

Fig. 1. Engagement, revision, and learning of AISF and AIPF, framed within SRL and SSRL.

the immediate task (Carless & Boud, 2018; Panadero, 2023). Focusing on processes of engagement, revision, and learning allows us to examine not only whether students improve their drafts (i.e., immediate revision performance) but also whether such improvement translates into long-term learning due to engagement. Their operation across AISF and AIPF contexts is summarized in Fig. 1.

Critically, engagement, revision, and learning are interdependent. Engagement with AI feedback shapes what kinds of revisions are made and what students learn (Fleckenstein et al., 2024). Further, revisions present an opportunity to solidify what was learned from feedback (Winstone & Carless, 2020), in part by strengthening retention through active learning. As Chi and Wylie (2014) noted in their ICAP (Interactive-Constructive-Active-Passive) engagement framework, revisions may activate comparing and contrasting prior knowledge, making next learning plans, or monitoring their understanding of tasks and materials. Specifically, revision may add a deeper understanding of how, when, and why certain actions should be performed through the process of translating received feedback into action (Bouwer & Dirkx, 2023; Cho & MacArthur, 2010). At the same time, there can also be a disconnect between revision and learning. When very detailed feedback is treated with low agency and reflection, improvements may be seen in revision but not in longer-term learning transfer (i.e., in future writing tasks; Wu & Schunn, 2021).

While individual engagement with GenAI often risks over-reliance (Fan et al., 2025), embedding these tools within collaborative learning environments can mitigate these pitfalls. Group settings trigger mechanisms of human-AI shared regulation in learning (Edwards et al., 2025), wherein peers proactively monitor, challenge, and justify one another's interactions with technology (Hadwin et al., 2018). This collective scrutiny transforms a potentially passive AI feedback reception task into a highly evaluative, shared cognitive process. In addition, the dual role of giving and receiving feedback that peers play (Wu & Schunn, 2023) may shape how students engage with AI feedback, the extent to which they revise, and what they learn. Notably, peers also serve as another cognitive resource in AIPF that is not found within AISF.

2.3. Research gap and the current study

Despite growing evidence on the effectiveness of GenAI-assisted feedback, prior studies have primarily examined improvements in immediate performance and revision outcomes (e.g., Guo et al., 2025; Lu et al., 2026). A small body of research has begun to offer process-oriented explanations of these gains, that is, by comparing improvement across tasks (Yan et al., 2026). However, these studies have typically focused on a single feedback context, and little is known about whether different AI-assisted feedback approaches (i.e., AISF and AIPF) lead to differences in students' engagement, immediate revision, and longer-term learning. Thus, we propose our first research question:

> RQ1. What are the relative effects of AISF and AIPF on engagement with AI, immediate revision performance, and longer-term learning transfer?

Our study also aims to examine how engagement and immediate revision performance are associated with longer-term learning transfer over time. Research in teacher, peer, or self-feedback contexts has shown that students' engagement with feedback is positively associated with writing performance (e.g., Jin et al., 2024; Lira Lorca et al., 2026). However, there is limited examination of how such gains are produced through revisions and engagement, particularly in AI-assisted feedback contexts. To address this gap, we propose our second research question:

> RQ2. Does engagement with AI feedback and immediate revision performance using AI-supported feedback similarly predict the amount of learning within AISF and AIPF?

3. Methods

This 16-week cluster-randomized experiment compared AISF and AIPF in terms of students' engagement, immediate revision performance, and longer-term learning transfer, and examined the learning processes that may account for the differences between the two conditions. This study is part of a larger project, with a prior manuscript examining intervention effects relative to different forms of teacher feedback.

3.1. Contexts and participants

Participants were 123 undergraduates enrolled in a compulsory English writing course at two technology institutions in central and eastern China. At each university, a single instructor taught multiple sections, and two of these sections participated in the current study. Students were majoring in Business English and had all passed the Chinese College English Test Band 4 (CET-4), corresponding approximately to a B1+ to B2 level on the Common European Framework of Reference for Languages (CEFR). Students were recruited through convenience and purposive sampling because their course teachers

participated in a professional development project focused on students' feedback engagement and writing development. The sample size was considered sufficient because the study employed a repeated-measures design with multiple drafts and revisions and included over 1100 writing samples and 432 reflection forms for analysis (see Appendix D for power analyses). Randomization was conducted at the class-section level, with intact sections at each university randomly assigned to either the AISF or AIPF condition. Ethical approval was obtained from the first author's institution. Informed consent was secured from both instructors and students, and participation was voluntary. All collected data were anonymized and used in accordance with ethical research guidelines. Across the semester, students completed ten writing tasks, including a pre-test, four formative tasks (each consisting of an initial draft and a revision), and a post-test. Of the initial samples, 110 students (AISF: $n = 51$; AIPF: $n = 59$) completed all four writing tasks and were included in the writing performance analysis (retention rate = 89.4%). In addition, 108 students (AISF: $n = 51$; AIPF: $n = 57$) completed all reflection forms and were included in the engagement analysis (retention rate = 87.8%).

3.2. Experimental design and procedures

Fig. 2 presents the cluster-randomized experimental design and study procedures.

Prior to the intervention, the researcher conducted two 60-min meetings with each course instructor to ensure consistency in instructional procedures and shared understanding of assessment criteria across conditions.

During the intervention, students in both conditions first completed a handwritten in-class pre-test writing task under teacher supervision. No external resources, including GenAI tools, were permitted to ensure that the task captured students' baseline writing proficiency. All students completed four formative writing tasks, each involving an initial draft followed by a revision. These repeated draft-revision cycles enabled the analysis of both immediate revision and subsequent writing over time. AISF and AIPF conditions differed in how AI was integrated into the feedback process, allowing for a comparison of how AISF and AIPF support immediate revision and longer-term learning. All writing tasks were argumentative because this genre aligns with instructional objectives and high-stakes test requirements. The writing topics differed across assignments to assess learning in novel contexts and reduce topic familiarity. All prompts were selected from previous TEM-4 (Test for English Majors-Band 4) examinations that students had not previously encountered. As standardized national English proficiency examinations for English majors in China, TEM-4 writing tasks follow a common standardized argumentative-writing format and are designed to assess a comparable level of task equivalence, topic difficulty, and prompt characteristics. Across tasks, the prompts required students to take a position, develop supporting reasons, and organize an argument using the same word-length requirements and assessment criteria. Each task required 250-300 words and was evaluated on idea relevance, organization, and language fluency. In both conditions, DeepSeek R1-0528 with DeepThink R1 function was used after students completed their drafts and during the revision stage. Throughout the intervention, role-assigned prompts embedded with writing rubrics and framed within Hattie and Timperley's (2007) feedback framework were employed.

In the AISF condition, students were scaffolded to actively elicit AI-generated feedback using role-assigned prompts (i.e., AI plays roles of an experienced teacher and a more proficient peer; see Appendix A). Students were encouraged to seek further feedback to address their specific learning needs, thereby supporting proactive engagement with AI feedback (Winstone et al., 2026). Following AI feedback interaction, students completed a structured self-reflection form (see Fig. 3a) guided by prompting questions, designed to facilitate evaluation of AI feedback

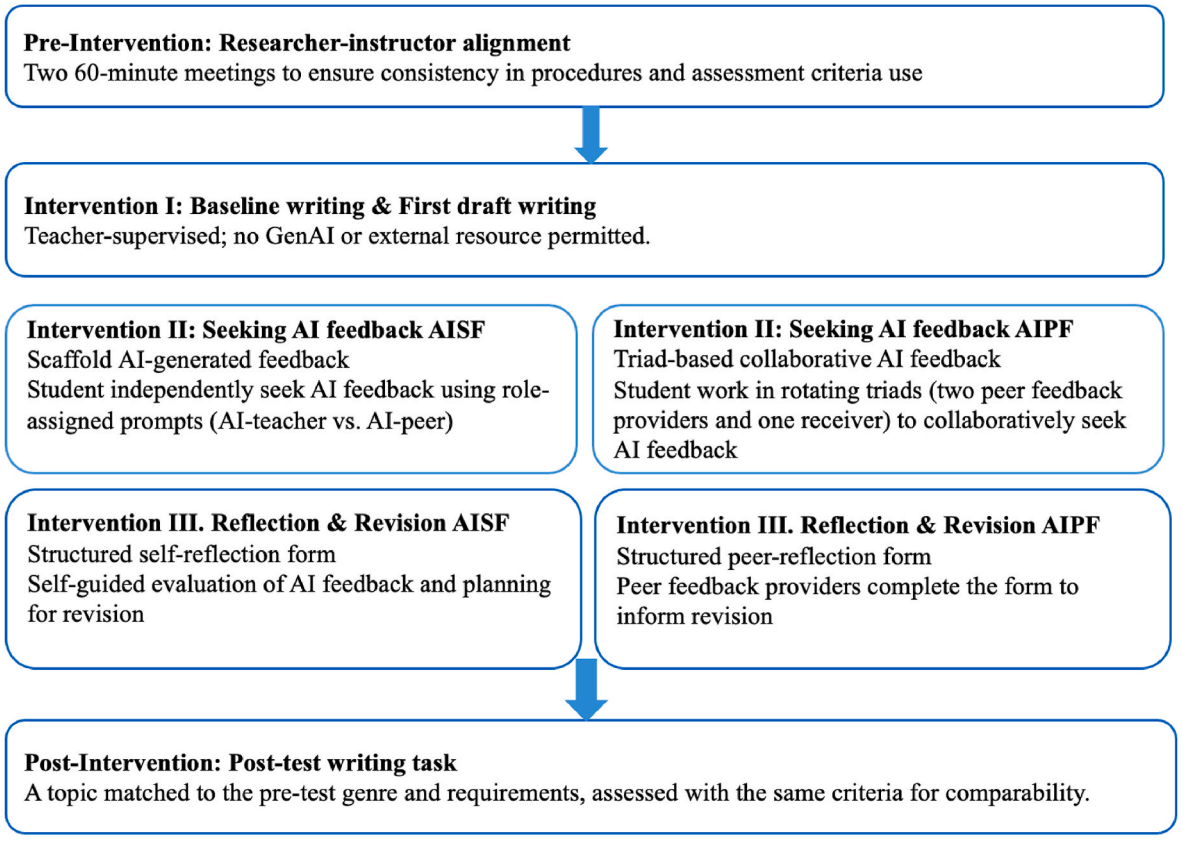


Fig. 2. Experimental design of AISF and AIPF procedures.

a) Self Reflection Form

1. When using different prompts in DeepSeek, what differences did you notice in the suggestions it provided?
2. Based on feedback from the AI teacher and AI peer, what do you see as the strengths of your essay? Please summarize them in terms of topic relevance, content, language, and structure.
3. Which suggestions from the AI-generated feedback did you adopt? Why did you choose to use these suggestions?
4. Based on feedback from the AI teacher and AI peer, how do you plan to adjust or improve your English learning before the next writing task?
5. Which suggestions from the AI teacher and AI peer do you think will remain useful for your next writing task?

b) Peer Reflection Form

1. Based on DeepSeek's feedback on your peer's essay and your subsequent discussion, what do you think are the strengths of your peer's essay? Please summarize them in terms of topic relevance, content, language, and structure.
2. After discussing DeepSeek's feedback, which revision suggestions did you recommend that your peer adopt? Why did you select these suggestions? Please explain in detail.
3. Based on DeepSeek's feedback and your own reflection, what adjustments or improvements could your peer make to their English learning before the next writing task?
4. Which suggestions from DeepSeek's feedback do you think will remain useful for your peer in their next writing task?

Fig. 3. Questions included in the reflection forms in the a) AI-Self and b) AI-Peer feedback conditions.

and planning for revision.

In the AIPF condition, students worked in triads consisting of two peer feedback providers and one feedback receiver. Group membership was assigned randomly by the instructor to avoid selection bias and to ensure varied peer interaction patterns, and roles were rotated across tasks so that each student equally served as a feedback giver and receiver during the intervention. Students as feedback providers first elicited AI feedback (see Appendix A), and then the triads collaboratively evaluated the AI-generated feedback and engaged in discussion to interpret its relevance and quality. Following the discussion, the peer feedback providers, led by the first peer, completed a structured peer reflection form (see Fig. 3b) to inform the following revisions.

After the intervention, all students in both groups undertook a post-test writing task. The post-test was designed to match the genre and task requirements of the pre-test and was evaluated using the same assessment criteria to ensure comparability.

3.3. Measures

3.3.1. Engagement in AISF and AIPF

Students' reflective forms were used as an indicator of their engagement across conditions, as well as a fidelity check of condition. These data were selected because they provide access to students' cognitive and metacognitive processes during feedback engagement, including feedback interpretation, evaluation, and higher-order reasoning. Our coding process combined deductive and inductive approaches. Deductively, the initial coding scheme was informed by feedback-related literature and theoretical frameworks, including evaluative judgement within feedback literacy (Carless & Boud, 2018) and models of GenAI feedback engagement (e.g., cognitive engagement; Zhan and Yan, 2026; self-regulated feedback engagement; Zhan et al., 2025). Inductively, additional codes were generated through close examination of the data to capture patterns not fully specified by existing theory, such as the extent to which feedback engagement aligned with students' learning needs. These codes were then refined and organized into categories through axial coding. Two researchers independently coded approximately 20% of the data, achieving substantial agreement (Cohen's $\kappa = 0.71$). Disagreements were resolved through discussion, after which the first author coded the remaining data.

Through an iterative coding process, a framework of five code rankings (1-5) was developed: no engagement = 1, passive adoption = 2, selection judgements = 3, strategic reasoning = 4, metacognitive or socially shared recognition = 5. A detailed coding scheme can be seen in Appendix B. Although AISF and AIPF involved different instructional activities by design, the coding framework was applied consistently across both conditions to assess the depth of cognitive engagement students demonstrated in response to AI feedback. The coding therefore focused on the depth of engagement using the same criteria across both conditions, irrespective of the instructional activity through which that engagement was elicited.

When a reflection contained indicators of more than one engagement level, it was assigned the highest level demonstrated. This decision reflected the subsumptive hierarchy underlying cognitive engagement, in which higher engagement modes build upon the processes presented by lower modes while indicating additional, more generative processing (Chi et al., 2018). It is also consistent with previous cognitive coding in which the highest level was applied when multiple levels were represented (Sailer et al., 2024). Thus, the score represents the most cognitively generative engagement demonstrated in the reflection rather than the frequency or average of all engagement indicators.

3.3.2. Writing and revision scores

Students' writing performance was assessed using the writing assessment criteria. Scores were rescaled from the original 0-15 range to

a 0-100 scale to align with formative assessment practices in both institutions. To ensure scoring consistency, all essays were evaluated using DeepSeek-V3.2 with a standardized prompt. Prior research has demonstrated that AI-based scoring can yield reliable and consistent evaluations with careful implementation and calibration, and it is advised to use it in low-stakes assessments (e.g., Kim & McCrudden, 2026). The AI-based scoring comprised two stages: a preparatory stage and a scoring stage. In the preparatory stage, a scoring prompt was developed by combining a role-assigned strategy with few-shot examples, both of which have been widely adopted in recent research (e.g., Pack et al., 2024). The prompt was initially drafted by the researchers, followed by iterative refinement with the assistance of ChatGPT4.0 and two writing instructors, and subsequently validated by two AI experts. An iterative refinement was conducted before full implementation.

In the scoring stage, a validation step was first conducted using essays from 62 randomly selected students ($n = 31$ for each university). These essays were independently scored by two expert raters and by the AI system to examine score comparability. The correlation between the mean AI scores and the mean expert scores was high ($rs = 0.84$, $ps < .001$), providing initial support for the validity of the AI-based scoring. This pilot phase also informed the refinement of the scoring prompt before formal assessment. Following the pilot, each essay (pre and post-test essays, along with both first and second drafts of each writing assignment) in the full dataset was evaluated three times using DeepSeek-V3.2, and the resulting scores were averaged to enhance reliability. The intraclass correlation coefficient (ICC) for the AI-generated scores was high (ICC = 0.97), indicating excellent consistency across repeated scoring.

Improvements in essays through revision from the first draft to the second draft for each assignment J were defined as the change in essay score: Δ_J = Draft 2_J – Draft 1_J.

3.4. Data analysis

For writing scores, each student contributes ten scores, including a pre-test, four formative tasks (each with an initial and revised draft), and a post-test. For engagement scores, each student contributes four scores across four tasks. The analytic sample comprised: (1) 1100 writing draft scores contributed by 110 students across two universities (University 1: $n = 520$; University 2: $n = 580$) and two conditions (AISF: $n = 510$; AIPF: $n = 590$) and (2) 432 engagement scores contributed by the same student cohorts (University 1: $n = 204$; University 2: $n = 228$).

3.4.1. Condition differences in engagement, revision, and learning

To examine whether the two conditions differed in engagement, revision, and learning outcomes (RQ1), three sets of analyses were conducted. First, differences in engagement were examined by comparing mean engagement scores in each assignment J (1–4) across conditions using descriptive statistics and mixed-effects modeling, with random intercepts for students nested within universities. Second, differences in revision were examined by modeling revision gains in each assignment J (1–4) as a function of condition and pre-test performance. A random intercept for students nested within universities was included to account for repeated measures. Third, differences in learning were examined using a lagged mixed-effects model in which, for each assignment J (1–3), subsequent first-draft performance (Draft1_{J+1}) was predicted from prior first-draft performance (Draft1_J) and condition. This model captures differences in writing performance across tasks prior to new feedback and revision, thereby reflecting learning from prior feedback and revision.

3.4.2. Associations of engagement and revision with learning transfer

To examine how engagement and revision contribute to learning (RQ2), mixed-effects models were estimated to capture relationships of 1) engagement to immediate revision performance, and 2) engagement and revision in a given task to longer-term learning transfer, operationalized as performance on the subsequent task. Specifically, to examine whether engagement predicts immediate revision performance, revised draft scores (Draft2_J) were modeled as a function of initial draft performance (Draft1_J), engagement, condition, and their interaction. Engagement was measured on a five-point scale and initially treated as a continuous predictor. This model tests whether greater engagement with AI feedback is associated with improved revision outcomes and whether this relationship differs across conditions. Follow-up analyses examined a theoretically meaningful distinction between lower (levels 1-3) and higher engagement (levels 4-5). To assess the robustness of this binary specification, supplementary analyses retained the original five-level engagement score and modeled it as either a continuous linear or continuous quadratic predictor, including its interaction with condition. The binary, continuous linear, and continuous quadratic specifications were estimated using maximum likelihood and compared using AIC and BIC. A likelihood-ratio test was also used to determine whether the quadratic specifications improved model fit relative to the continuous linear specification. Likelihood-ratio results are reported in Appendix F.

To examine whether revision predicts learning, a lagged mixed-effects model was estimated in which subsequent first-draft performance (Draft1_{J+1}) was predicted from prior first-draft performance (Draft1_J), revision gain (Δ_J), condition, and the interaction of condition with revision gain. By controlling for prior performance, this model tests whether improvements achieved through revision are associated with better performance on subsequent tasks.

To examine the joint contribution of revision and engagement (beyond its direct effects on revision) to learning, an additional mixed-effects model was estimated in which subsequent first-draft performance (Draft1_{J+1}) was predicted from prior first-draft performance (Draft1_J), engagement in assignment J, revision gain (Δ_J), condition, and the interactions of condition with engagement and revision. Engagement was modeled as binary, reflecting an examination of the relationship of each engagement level to the learning outcome. Assignment index was included as a control variable, and random intercepts for students and universities were specified to account for repeated measures and clustering. This model allows for the simultaneous examination of (a) the association between revision gains and longer-term learning transfer, (b) the direct contribution of engagement to longer-term learning transfer, and (c) condition differences in these relationships. Cohen's d was calculated for all comparisons to quantify the magnitude of the differences. Effect sizes were interpreted as small ($d \approx 0.2$), medium ($d \approx 0.5$), and large ($d \approx 0.8$) (Cohen, 1988).

4. Findings

4.1. Condition differences in engagement, revision, and learning (RQ1)

We first examined whether the AISF and AIPF conditions differed in engagement, immediate revision performance, and longer-term learning transfer (see Fig. 4).

Students in the AIPF condition reported slightly higher levels of engagement with AI feedback ($M = 3.37$, $SD = 0.96$) than those in the AISF condition ($M = 2.97$, $SD = 1.36$), $F(1, 106) = 8.92$, $p = .004$, Cohen's $d = 0.35$. The proportion of higher engagement (i.e., 4s or 5s) was also slightly higher in the AIPF (42%) than in AISF (40%), but this difference was not statistically significant, $F(1, 430) = 0.26$, $p = .61$, Cohen's $d = 0.05$. Across all reflections, engagement was moderate rather than uniformly intensive($M = 3.18$, $SD = 1.18$, median = 3), with higher-level engagement occurring in 41% of tasks. This pattern

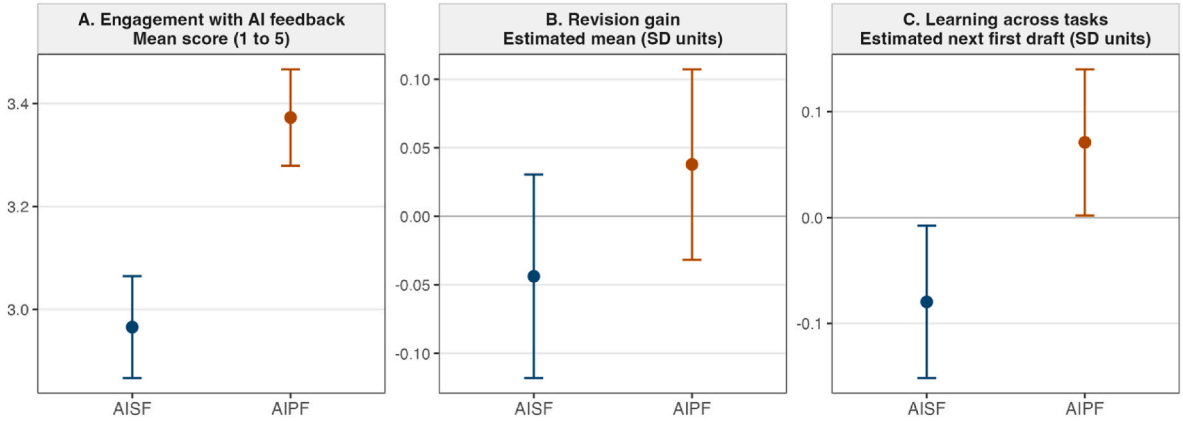


Fig. 4. Estimated means (with SE bars) for (A) engagement with AI feedback, (B) revision gains, (C) subsequent first-draft performance by condition.

suggests that higher-level engagement occurred in a substantial minority of tasks, although lower levels of engagement remained more common overall. Revision gains averaged 7.12 points ($SD = 4.51$, median = 6.83), and most task-level observations showed positive gains in both AISF (94.6%) and AIPF (96.6%) (see Appendix H).

Revision gains, defined as the difference between revised and initial drafts, did not differ significantly between conditions ($\beta = 0.08$, 95% CI [−0.12, 0.28], $p = .42$), indicating that students in AISF and AIPF improved their drafts to a similar extent within each task. In contrast, condition differences were observed in longer-term learning transfer across tasks. Students in the AIPF condition demonstrated substantially higher subsequent first-draft performance compared to those in AISF ($\beta = 0.30$, 95% CI [0.08, 0.51], $p = .006$), after controlling for prior performance.

These findings indicate that AIPF students exhibited slightly higher engagement and significantly better performance across tasks, although this advantage was not accompanied by greater immediate revision gains. The advantage of AIPF therefore emerged in longer-term learning transfer rather than in immediate revision performance.

4.2. *Engagement and revision as mechanism of learning (RQ2)*

In this section, we examined how engagement and immediate revision contribute to longer-term learning across tasks.

4.2.1. *Engagement and revision*

The model results showed that engagement with AI was a significant positive predictor of immediate revision performance ($\beta = 0.09$, 95% CI [0.01, 0.16], $p = .024$), indicating that greater engagement with AI feedback was associated with improved revision outcomes. However, this relationship differed by condition. The interaction between engagement and condition was significant and negative ($\beta = -0.13$, 95% CI [−0.26, −0.01], $p = .034$), indicating that the positive effect of engagement on revision was weaker in the AIPF condition. Follow-up analyses used the binary high/low specification of engagement because the relationship with revision appeared to follow a threshold-like pattern rather than a linear or quadratic pattern (see Appendix C and F). Thus, higher engagement was associated with greater immediate revision gains, but this advantage was limited to only the AISF condition (see Fig. 5).

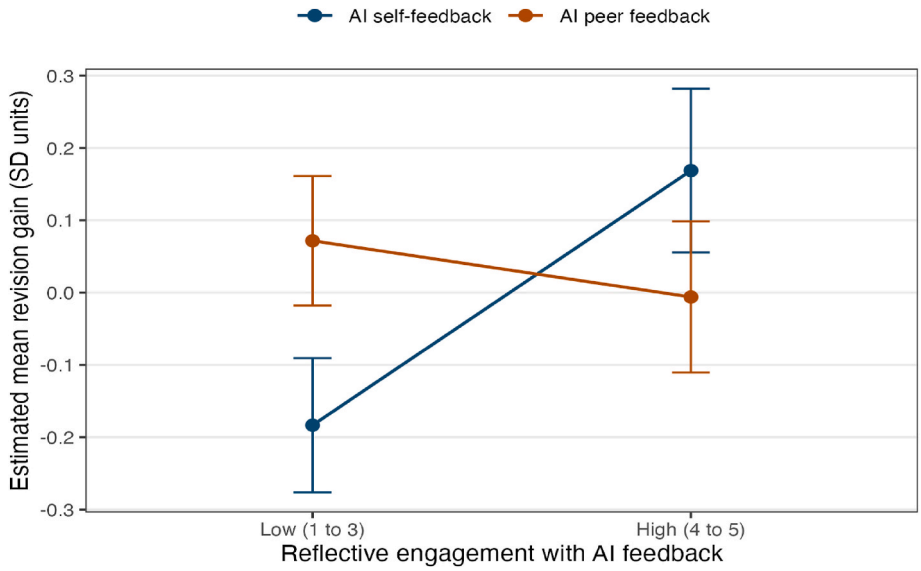


Fig. 5. Mean estimated revision gains (with SE bars) by Low vs. High engagement with AI feedback within each condition.

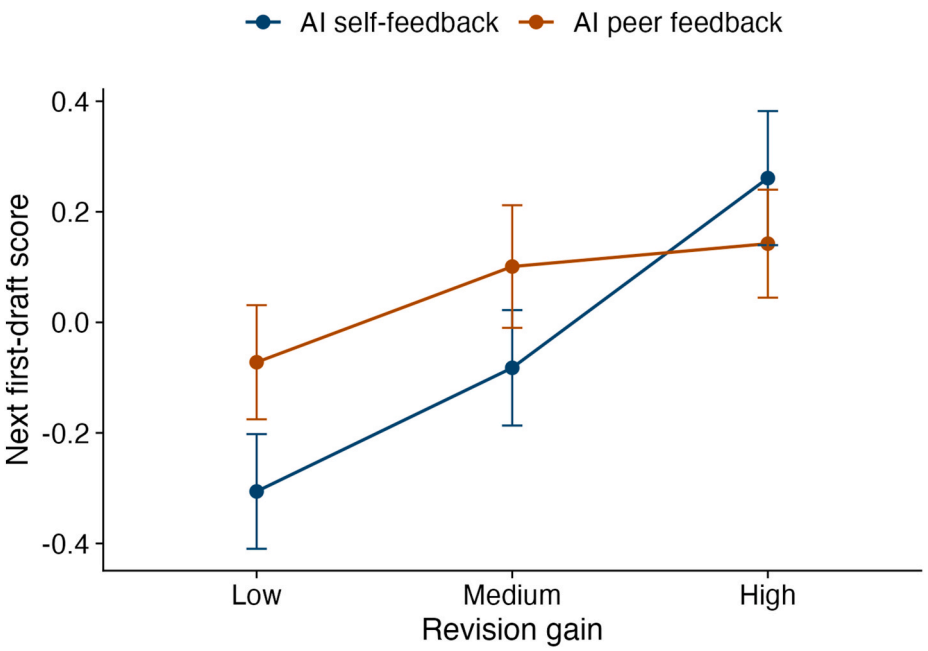


Fig. 6. Mean estimated next assignment first draft score as a function of revision level and feedback condition.

4.2.2. Revision and learning

Revision gains were a significant predictor of subsequent first-draft performance ($\beta = 0.32$, 95% CI [0.20, 0.44], $p < .001$), suggesting that stronger immediate revision performance was associated with greater longer-term learning transfer. However, this relationship was also strongly moderated by condition ($\beta = -0.28$, 95% CI [−0.44, −0.11], $p = .001$). As shown in Fig. 6, students in the AISF condition appear to learn more when they implement substantial immediate revisions, whereas students in the AIPF condition appear to show a higher level of learning transfer across different levels of revision gains.

4.2.3. Integrated model of learning

To examine the joint contributions of engagement and immediate revision, an integrated model was estimated in which subsequent first-draft writing performance was predicted from prior performance, engagement, revision gains, condition, and their interactions. The results showed that both engagement with AI ($\beta = 0.26$, 95% CI [0.01, 0.52], $p = .041$) and revision gains ($\beta = 0.29$, 95% CI [0.17, 0.40], $p < .001$) were significant predictors of subsequent writing performance. In addition, condition remained a significant predictor ($\beta = 0.26$, 95% CI [0.04, 0.47], $p = .018$), indicating that AIPF students outperformed AISF

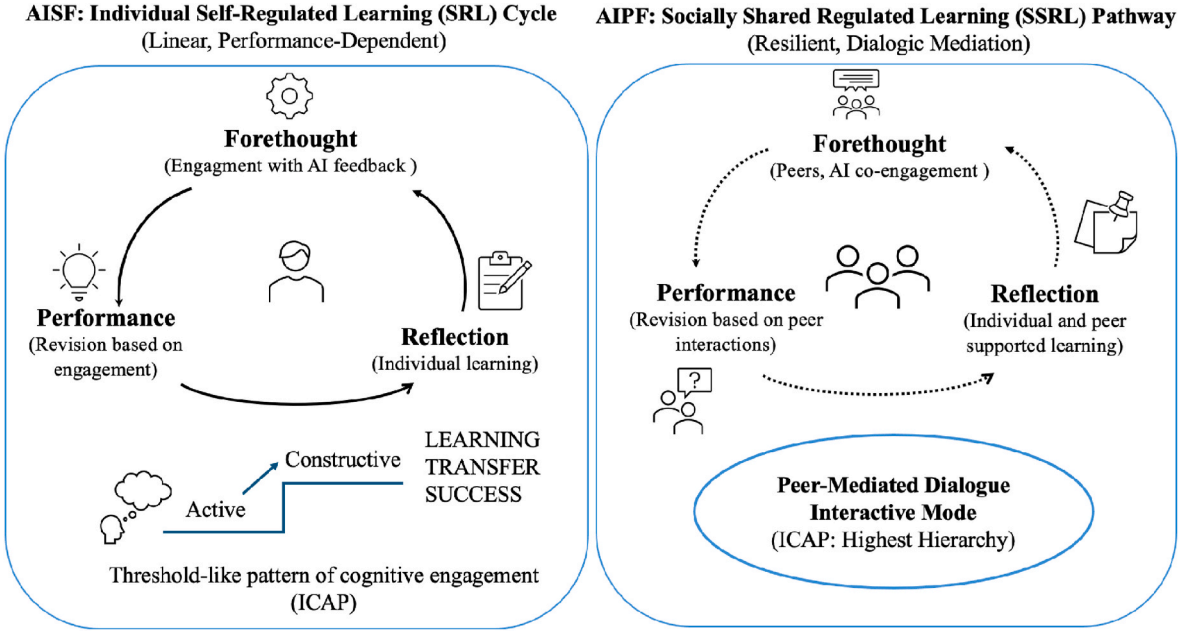


Fig. 7. SRL and SSRL learning pathways within each condition.

students even after accounting for engagement with AI and revision. The interaction between revision and condition was negative and significant ($\beta = -0.25$, $p = .002$), suggesting that the contribution of immediate revision to longer-term learning was weaker in AIPF. The interaction between engagement and condition was also negative and similar to the size of the main effect, but not statistically significant ($\beta = -0.28$, $p = .105$).

These findings indicate that engagement with AI and immediate revision gains were independently associated with longer-term learning transfer in the AISF condition. In contrast, the associations were weaker in the AIPF condition, despite students demonstrating stronger overall learning transfer across tasks.

5. Discussion

5.1. Interpretation of findings

5.1.1. Effects of AI-assisted feedback contexts on engagement, revision, and learning

Our findings for RQ1 align with a process-oriented feedback perspective (Carless & Boud, 2018; Winstone et al., 2017) in which engagement with feedback, immediate revision, and longer-term learning are importantly different. The observed differences in engagement and revision across the AISF and AIPF conditions extend earlier research that primarily focused on a single mode of AI-assisted feedback (e.g., Yan et al., 2026). Likewise, the finding that AIPF yields higher learning gains than AISF extends previous research on comparisons between peer feedback and self-feedback from non-AI contexts (e.g., Heil & Ifenthaler, 2026; Zou et al., 2023). Our findings particularly highlight the importance of tracking impacts on longer-term learning rather than only focusing on immediate revision behaviors or revision quality, which we discuss in greater depth when interpreting the observed learning processes.

5.1.2. Learning pathway in AISF and AIPF

For AISF, our findings indicate that engagement was a positive predictor of immediate revision, which in turn predicted longer-term learning transfer. This is consistent with evidence that engagement with LLM feedback supports revision quality (Glüsing et al., 2026) and that revision following feedback can improve performance on subsequent writing tasks (Ekanayaka & Ellis, 2020). From an SRL perspective, individual learning operates through a path of forethought, performance, and reflection (see Fig. 7). However, this pathway appears to rely more heavily on the depth of individual cognitive engagement, exposing a clear threshold-like pattern: learning gains emerge only when engagement reaches a sufficient depth. Without external social accountability, individual students in the AISF condition risk cognitive passivity or mere active engagement (Chi & Wylie, 2014), where they simply copy or restate AI-generated feedback without deep processing, leading to their overreliance on GenAI (Fan et al., 2025). For these individual learners, engagement only reaches a truly constructive level if they possess the self-regulatory capacity to reflect on their performance, monitor their understanding of AI-generated feedback, and make revision plans to close the feedback loop (Zhan et al., 2025).

For AIPF, a different pathway emerged, where the relationship of longer-term learning transfer to both AI engagement and immediate revision performance was markedly weaker. In this context, students demonstrated superior learning transfer regardless of their individual revision quality or engagement levels. From an SSRL perspective, the instructional design created opportunities for shared monitoring, collective processing, and social interactions that may have contributed to the longer-term learning transfer observed. These regulatory processes were enacted through concrete learning activities such as jointly evaluating AI-generated feedback, articulating rationales for feedback use, discussing revisions, and elaborating on alternative interpretations. Such activities have been shown to promote learning (e.g., Nicol et al., 2014; Wu & Schunn, 2023) and may represent important processes through which socially shared regulation operates. Collectively, these collaborative learning activities may create a socially accountable learning environment that extends beyond individual engagement with AI feedback.

The social accountability may mitigate individual cognitive offloading when interacting with GenAI and promote more constructive learning processes, such as AI feedback prompting and discussion with peers about AI feedback interpretations. Through these collaboration processes, students may develop a deeper understanding of writing quality and transfer this understanding to subsequent writing tasks, even when their immediate task-level revisions were modest. Our findings are therefore consistent with previous research showing the importance of peer dialogic interactions for learning in non-AI-mediated learning environments (e.g., Er et al., 2021; Lira Lorca et al., 2026).

Our findings are also consistent with prior findings showing that growth in performance from peer feedback depends upon the depth of engagement in providing feedback (Hornstein et al., 2025; Yu & Schunn, 2025). Within the ICAP engagement framework, AIPF naturally facilitates interactive engagement because learners engage in dialogic exchanges about the quality, relevance, and usefulness of AI feedback. Such interactions may create conditions for "co-inferring" (Chi & Wylie, 2014, p. 226) when each party creates new knowledge and perspectives in a cumulative and spiraling manner. One possible explanation is that, within the AIPF condition, a peer feedback giver evaluates AI-generated feedback and shares their interpretation with peers, whose responses (e.g., accepting, rejecting, justifying, reasoning) further extend or challenge this interpretation. When conflicting views arise, further rounds of meaning negotiation are triggered, leading learners to collaboratively construct understandings of writing quality and assessment criteria that can be internalized and transferred to future tasks (Nicol et al., 2014; Topping et al., 2025; Wu & Schunn, 2021), even if such learning may not be immediately observed (Little et al., 2025; Zhang et al., 2023).

5.2. Contributions, implications, and limitations

5.2.1. Theoretical contributions

This study makes three theoretical contributions. First, it extends theories of self-regulated learning (SRL) and socially shared regulated learning (SSRL) into AI-assisted feedback contexts by identifying different patterns of engagement and learning across AISF and AIPF. The pattern observed in AISF was consistent with a more linear, performance-dependent SRL process, in which learning transfer was associated with students' cognitive engagement and immediate revision gains. In AIPF, stronger learning transfer was observed even when these associations were weaker. This pattern is consistent with opportunities for socially shared regulation. However, it cannot be attributed exclusively to SSRL, as it may also reflect related collaborative learning processes (e.g., peer explanation, dialogic interaction, elaborative processing) that were not explicitly measured in the current study. These findings suggest that social interaction may transform AI from a correction tool into a catalyst for collaborative metacognition in AI-assisted peer feedback.

Second, this study provides a more fine-grained understanding of cognitive GenAI feedback engagement through the ICAP framework (Chi & Wylie, 2014). Our findings suggest that merely 'actively' engaging is insufficient in AISF contexts; instead, learning requires constructive engagement through generating new ideas and

interpretations. By contrast, AIPF provides opportunities for interactive engagement, which is the highest mode in the ICAP hierarchy. By highlighting the role of social interactions in co-creating knowledge and interpretations, this study suggests that AI feedback may be most effective in the long term when it fosters interactive meaning-making and co-construction of understanding. Consequently, we advocate that scholars differentiate levels of engagement (e.g., active, constructive, interactive) rather than using the broader term 'active' engagement (cf. Chi et al., 2018).

Finally, the AISF and AIPF designs of this study contribute to the literature on a dialogic view of feedback (e.g., Carless & Boud, 2018; Winstone & Carless, 2020) by elucidating how students evaluate, understand, seek, and use AI feedback throughout their learning process (Chen & To, 2026).

5.2.2. Implications, limitations, and future research

More generally, teachers can use AI not merely as a direct feedback generator but a cognitive stimulus for expanding feedback provision and quality evaluation, or as a dialogic resource that supports students' reflection on writing quality (Chen et al., 2026). Those AI-mediated feedback activities should particularly be embedded within structured draft-redraft cycles (Panadero, 2023; Winstone & Carless, 2020) that explicitly connect AI feedback engagement to revision and subsequent performance. In particular, teachers can design AI-assisted feedback activities that foster deeper levels of engagement (constructive or interactive) to support durable performance. This was particularly important for learning from AISF. Although such engagement was not critical for learning from AIPF, there may still be important aspects of eliciting useful feedback. For example, AI within peer feedback may help introduce alternative ideas, perspectives, and feedback into peer discussions, thereby extending students' reflection prior to revision.

From a design perspective, AI systems could be developed to detect and respond to learners' engagement levels, providing more structured scaffolding when students' engagement remains at a surface level. Building on the benefits of AIPF identified in the current study, AI agents also hold promise for fostering socially shared regulation or human-AI shared regulation (Edwards et al., 2025).

We acknowledge that the retention interval in this study is relatively short (one week). Future research should employ delayed follow-up tests or transfer tasks across extended intervals, such as in a subsequent semester, to evaluate whether the superior learning transfer observed in the AIPF remains durable over time once all scaffolding is removed. Our findings suggest the value of integrating SRL and SSRL perspectives with cognitive engagement frameworks (e.g., ICAP) to better explain how human-AI collaborative processes operate across contexts. Because the present intervention differed simultaneously in peer interaction, collaborative discussion, accountability, and opportunities for shared regulation, future studies should experimentally disentangle the relative contribution of these collaborative learning mechanisms.

Several other limitations of this study should be acknowledged. First, a key limitation concerns the interpretation of the underlying learning mechanisms. The two conditions differed in several instructional features beyond opportunities for regulation. Specifically, the AIPF group includes peer explanation, collaborative discussion, mutual accountability, and feedback generation processes. Consequently, although the findings are consistent with the SSRL perspective, the present design does not allow the contribution of socially shared regulation to be disentangled from other collaborative learning mechanisms, such as jointly evaluating AI-generated feedback, discussing revisions, or articulating feedback rationales. Future research could isolate these processes using more controlled experimental designs or fine-grained process analyses. Second, this study was conducted within a specific group and domain. Although the repeated-measures design provided substantial observations across tasks, the observed findings remain situated within a specific disciplinary, educational, and cultural context. The generalizability of findings to other disciplines, age groups, educational levels, and cultural settings therefore requires further investigation. Future research could examine whether similar patterns emerge in different subject areas (e.g., STEM education), learner populations (K-12 students), and institutional and cultural contexts (e.g., vocational colleges or universities outside China). Third, engagement was measured through students' reflective forms, which primarily capture cognitive and metacognitive aspects of engagement. While this approach provides insights into learners' thinking processes, it may not fully capture the dynamic processes through which students interact with AI feedback and peers during learning. Future research can integrate richer process measures, such as AI interaction logs, prompt histories, revision traces, and transcripts of peer discussions, together with qualitative discourse analysis to provide a more comprehensive understanding of the cognitive and regulatory processes underlying AI-assisted feedback processes. In addition, engagement was operationalized as the highest level of cognitive engagement demonstrated in students' reflections on AI feedback. Although this approach captured the most cognitively generative processing demonstrated by each student, it did not account for the frequency or consistency of different engagement levels across reflections. Future research could employ frequency-based or longitudinal measures to examine how the consistency and temporal patterns of engagement relate to learning outcomes.

Fourth, the study investigated two structured AI-assisted feedback designs without comparison with more naturalistic conditions. The findings should therefore be interpreted as applying to scaffolded AI use supported by carefully designed meta-prompts. Future research could compare AISF and AIPF with conventional feedback approaches (e.g., self-feedback, peer feedback, or teacher feedback) and less scaffolded GenAI feedback to examine differences in students' engagement, immediate revision performance, and longer-term learning transfer. Finally, using AI in both the intervention and assessment may have introduced shared-method effects. However, the two procedures used different model versions, prompts, and purposes, and AI scores showed strong agreement with expert ratings. Although these safeguards reduce this concern, subtle alignment cannot be excluded. Future research could use human-scored outcomes independently or an assessment model from a different model family.

6. Conclusion

This study compared the effects of AI-assisted self-feedback and AI-assisted peer feedback on engagement, revision, and learning within a 16-week cluster-randomized experiment. The findings suggest different patterns of engagement and learning across two instructional conditions that are consistent with SRL and SSRL perspectives. Specifically, AISF appeared more dependent on students' constructive engagement with AI feedback and the quality of their revisions, whereas AIPF was associated with stronger learning transfer in a manner consistent with opportunities for socially shared regulation and collaborative engagement.

CRediT authorship contribution statement

Shijun (Cindy) Chen: Writing – original draft, Visualization, Project administration, Methodology, Investigation, Data curation, Conceptualization, Resources, Writing – review & editing. **Conrad Borchers:** Formal analysis, Methodology, Visualization, Writing – review & editing, Writing – original draft. **Christian D. Schunn:** Supervision, Methodology, Conceptualization, Validation, Writing – review & editing.

Appendix A. Prompts Used in AISF and AIPF Conditions

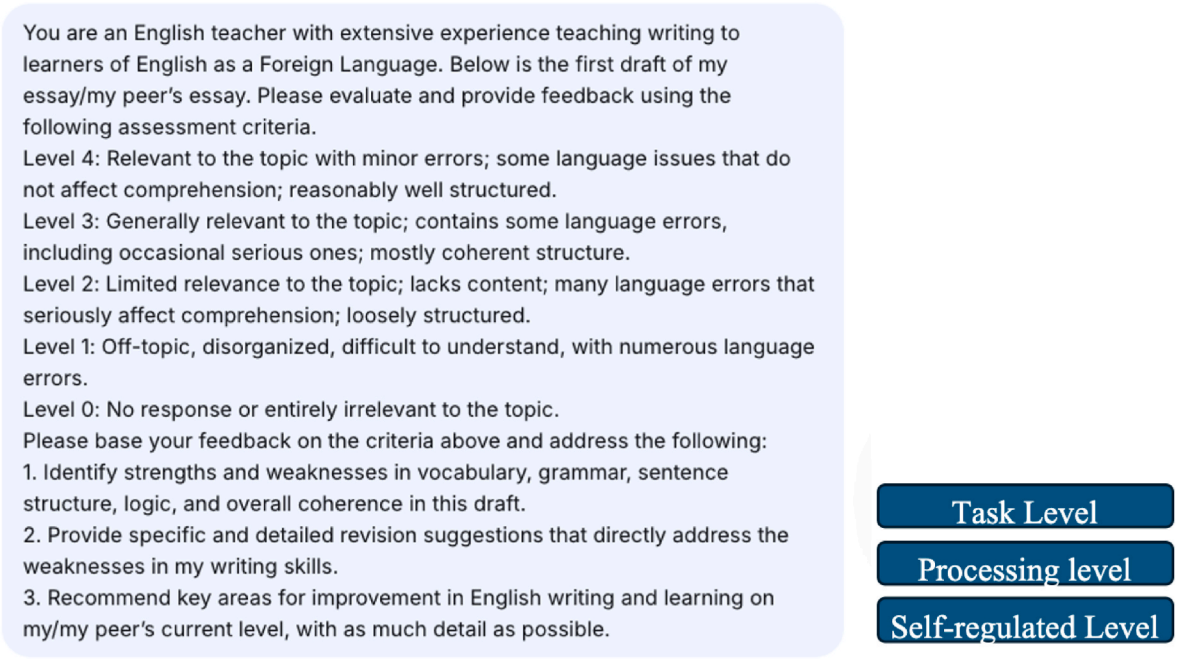


Fig. A1. Role-assigned prompt with embedded assessment criteria.

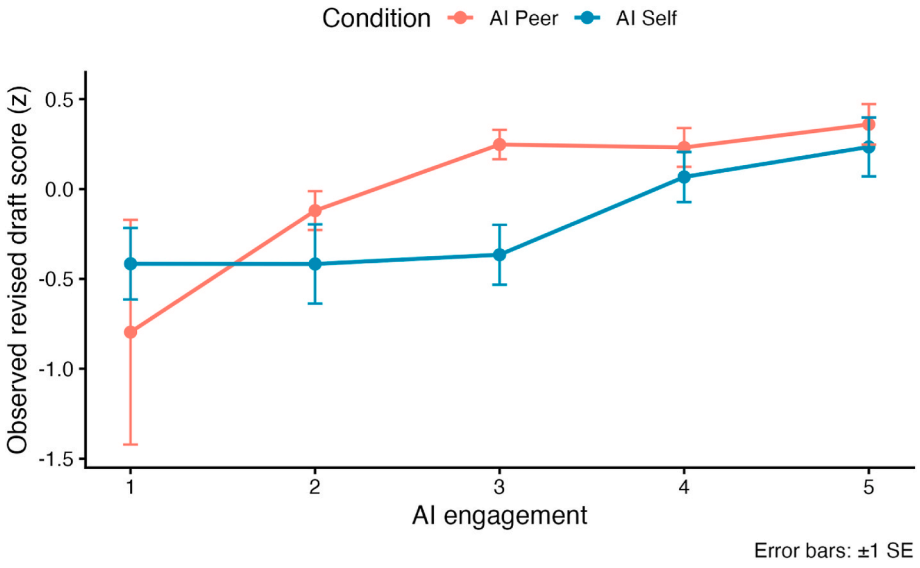
1. I think the word "very" is too basic. Based on the meaning and context of my writing, please suggest more precise and appropriate adverbs to replace it.
2. I think the following sentence [paste your sentence] contains some grammatical errors, but I'm not sure what they are. Please identify and explain these errors and provide some relevant grammar exercises to help me practise.
3. Here is my English essay [paste your essay or an excerpt]. I think it includes some Chinglish expressions. Please identify these expressions and suggest more natural, idiomatic alternatives.
4. Here is my English essay [paste your essay or an excerpt]. I think the logical connection between the first and second sentences is weak. Please explain why the transition may be unclear and suggest ways to improve the coherence between them.
5. Here is my English essay [paste your essay or an excerpt]. I think the examples I've provided are insufficient to support my argument. Based on my thesis statement, please suggest stronger or more varied examples.

Fig. A2. Sample prompts for seeking AI-generated feedback. (Notes: The main prompt at the top, with annotation for which aspects of the prompt correspond to the task, process, and self-regulated levels). "my essay/my peer's essay" is the content that varies by condition. Suggested prompt variations are shown below, where "my essay" was replaced with "my peer's essay" in the AIPF condition.).

Appendix B. Engagement with AI Feedback Coding Scheme

Pattern	Definition	AISF Example	AIPF Example
No engagement 1	Individual does not address or incorporate AI feedback in the reflection.	N/A	N/A
Passive adoption 2	The individual accepts AI feedback as is, whether surface-level or content-level, without questioning its accuracy, relevance, or appropriateness for the task or context.	Rewrites sentences or changes arguments exactly as AI suggests, without evaluating fit or correctness.	
Selective judgements 3	The individual reviews multiple AI feedback points and selects which ones to apply, showing some judgement, but lacks sound explanation or justification for his or her choices.	Applies some AI suggestions but ignores others, based on perceived usefulness, without sound reasoning.	
Strategic reasoning 4	The individual explicitly explains the rationale behind accepting or rejecting specific AI feedback, showing a thoughtful decision-making process.	AI feedback: There should be an addition of counterarguments discussing the negative effects of excessively restricting screen time. Self-feedback: I think by adding counterarguments, the essay becomes more comprehensive and dialectical, thereby enhancing its persuasiveness.	AI feedback: change show ideas into voice perspectives Peer reviewers: Rejected. "Show ideas" did not occur in my peer's essay.
Meta-cognitive recognition 5	Individual articulates awareness of learning gaps and plans for improvement.	AI feedback: Use IELTS materials which can be directly transferred to TEM-4; Search for social issues in Google Scholar to access cutting-edge research papers. Self-feedback: The advice is not very suitable for me in TEM-4 test preparations. IELTS and academic papers are beyond my language abilities.	AI feedback: The essay can follow the structure of "problem—reason—solutions" Peer reviewers: We think this AI feedback is practical for our peers and can also be applied in his future essays.

Appendix C. Marginal Revised Draft Scores as a Function of Detailed AI Engagement Levels in AISF and AIPF Conditions, Controlling for Draft 1 Score



Appendix D. Power Analyses

A priori power analyses were conducted in *R* to evaluate whether the planned analytic sample provided adequate sensitivity for the RQ1 and RQ2 mixed-effects models. For writing scores, the analysis drew on data from 110 students, comprising 440 revision observations, and 330 lagged observations. For feedback engagement, the analysis drew on data from 108 students, comprising 432 task-level engagement and revision observations, and 324 lagged learning observations. Power was estimated for benchmark standardized effects of $d = 0.30$, 0.40, and 0.50 drawn from prior meta-analytic work on AI-assisted feedback (Liu & Liang, 2026). Across 10,000 simulation iterations per model term and benchmark effect, power for RQ1 condition effects and RQ2 main effects and interactions generally ranged from 73% to 100% for $d = 0.30$, 93% to 100% for $d = 0.40$, and 97% to 100% for $d = 0.50$, indicating adequate sensitivity for small-to-moderate effects. The only exception was the high-engagement by condition interaction in the integrated learning model, which showed lower sensitivity, with power estimates of 41% for $d = 0.30$, 63% for $d = 0.40$, and 82% for $d = 0.50$, corresponding to a minimum detectable effect of $d = 0.49$ for 80% power. A two-sided alpha level of .05 was used throughout.

Appendix E. Tables Correlation Matrix

Table E1
Correlation matrix for revised-score model variables.

Variable	1	2	3	4	5	6	7
1. university_F	1.00						
2. condition_ai_peer	0.02	1.00					
3. draft_num	0.00	0.00	1.00				
4. draft1_z	−0.27	0.15	0.59	1.00			
5. draft2_z	−0.32	0.19	0.45	0.78	1.00		
6. engagement_raw	−0.03	0.17	0.32	0.22	0.22	1.00	
7. engagement_high	0.02	0.02	0.28	0.17	0.17	0.84	1.00

Table E2
Correlation matrix for next first-draft model variables.

Variable	1	2	3	4	5	6	7
1. university_F	1.00						
2. condition_ai_peer	0.02	1.00					
3. draft_num	0.00	0.00	1.00				
4. draft1_prev_z	−0.31	0.15	0.60	1.00			
5. draft1_next_z	−0.21	0.17	0.43	0.64	1.00		
6. engagement_raw	−0.02	0.18	0.27	0.19	0.20	1.00	
7. engagement_high	0.02	0.00	0.24	0.14	0.17	0.83	1.00

Table E3
Correlation matrix for student-level source variables.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. university_2	1.00												
2. condition_ai_peer	0.02	1.00											
3. draft 1_raw	0.63	0.12	1.00										
4. draft 2_raw	0.40	0.21	0.52	1.00									
5. draft 3_raw	0.07	0.16	0.16	0.40	1.00								
6. draft 4_raw	0.25	0.22	0.27	0.45	0.65	1.00							
7. draft 5_raw	0.15	0.11	0.41	0.26	0.16	0.28	1.00						
8. engagement_task1	0.03	0.18	0.07	0.03	−0.04	−0.01	0.03	1.00					
9. engagement_task2	0.13	0.15	0.02	0.08	0.03	0.10	0.03	0.35	1.00				
10. engagement_task3	−0.23	0.23	0.26	0.28	0.19	0.24	0.21	0.19	0.23	1.00			
11. engagement_task4	−0.03	0.17	0.06	0.11	0.12	0.03	0.14	0.28	0.20	0.17	1.00		
12. engagement_mean	−0.04	0.28	0.10	0.19	0.11	0.14	0.16	0.70	0.68	0.62	0.62	1.00	
13. engagement_high_prop	0.04	0.04	0.03	0.12	0.09	0.08	0.07	0.55	0.65	0.49	0.51	0.84	1.00

Appendix F. Supplementary analysis of engagement specification

Table F1
Model-fit comparison across engagement specifications predicting revision.

Engagement specification	AIC	BIC	Log likelihood	ΔAIC	ΔBIC
Binary: levels 1–3 vs. 4–5	871.74	908.36	−426.87	0.00	0.00
Continuous linear	874.82	911.44	−428.41	3.08	3.08
Continuous quadratic	877.76	922.51	−427.88	6.02	14.15

Appendix G. Scoring prompt

You are a university English teacher. Please evaluate the following student essay according to the rubric below. The score ranges from 0 to 100. Your scoring must be objective, consistent, and reproducible. Use consistent standards across different essays.

Rubric:

90–100 (Excellent): Highly clear, relevant, and insightful; strong logical arguments with originality; excellent organization; fluent and accurate

language with almost no errors.

80–89 (Very Good): Clear and relevant; strong arguments; well-organized with minor weaknesses; language mostly accurate with minor errors.

70–79 (Good): Mostly relevant with clear ideas; reasonable but less original arguments; acceptable organization with weaknesses; language has some errors but meaning is clear.

60–69 (Fair): Somewhat relevant but limited ideas; weak argumentation; uneven organization; language errors sometimes affect clarity.

50–59 (Poor): Limited relevance; underdeveloped or illogical arguments; weak organization; frequent language errors affecting clarity.

0–49 (Very Poor): Off-topic, unclear, or underdeveloped; incoherent arguments; missing organization; language errors make comprehension very difficult.

Appendix H. Descriptive Statistics and Distributions of Engagement and Revision Gains

Table H1
Descriptive statistics for engagement and revision gains by condition

Measure	AISF	AIPF	Overall
Engagement, M (SD)	2.97 (1.36)	3.37 (0.96)	3.18 (1.18)
Engagement, median	3	3	3
Lower engagement, levels 1–3, n (%)	123 (60.3%)	132 (57.9%)	255(59%)
Higher engagement, levels 4–5, n (%)	81 (39.7%)	96 (42.1%)	177(41%)
Revision gain, M (SD)	6.92 (4.53)	7.29 (4.50)	7.12 (4.51)
Revision gain, median (IQR)	6.33 (4.00-9.33)	7.00 (4.67-10.00)	6.83 (4.33-9.67)
Positive revision gain, n (%)	193 (94.6%)	228 (96.6%)	421 (95.7%)
No revision gain, n (%)	2 (1%)	1 (0.4%)	3 (0.7%)
Negative revision gain, n (%)	9 (4.4%)	7 (3.0%)	16 (3.6%)

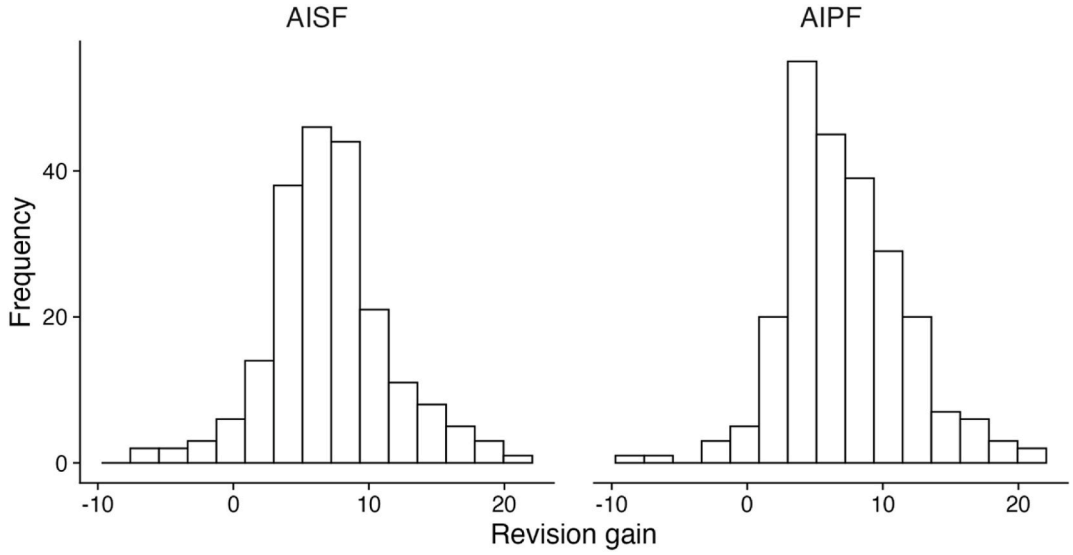


Fig. H1. Distribution of immediate revision gains by condition.

Data availability

The data that has been used is confidential.

References

Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education: Principles, Policy & Practice, 27*(4), 332–349. https://doi.org/10.1080/0969594X.2019.1609411

Bouwer, R., & Dirkx, K. (2023). The eye-mind of processing written feedback: Unraveling how students read and use feedback for revision. *Learning and Instruction, 85*, Article 101745. https://doi.org/10.1016/j.learninstruc.2023.101745

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245–281. https://doi.org/10.2307/1170684

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

Chen, S.(C.), & Carless, D. (2026). Teacher feedback dilemmas and the use of GenAI: Challenges or opportunities? *Innovations in Education & Teaching International*, 1–14. https://doi.org/10.1080/14703297.2026.2648040

Chen, S.(C.), & To, J. (2026). Unravelling undergraduates' development of evaluative judgments through AI-supported internal feedback. *Assessment & Evaluation in Higher Education*, 1–18. https://doi.org/10.1080/02602938.2026.2711099

Chen, S.(C.), Zhang, L., & Li, A. W. (2026). Leveraging self-regulated learning to support university students' feedback literacy in blended learning environments. *Assessment & Evaluation in Higher Education, 51*(1), 174–193. https://doi.org/10.1080/02602938.2025.2546349

Chi, M. T. H., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., Levy, R., Li, N., McEldoon, K. L., Stump, G. S., Wylie, R., Xu, D., & Yaghmourian, D. L. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science, 42*(6), 1777–1832. https://doi.org/10.1111/cogs.12626

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328–338. https://doi.org/10.1016/j.learninstruc.2009.08.006

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). L. Erlbaum Associates.

Corbin, T., Tai, J., & Flenady, G. (2025). Understanding the place and value of GenAI feedback: A recognition-based framework. *Assessment & Evaluation in Higher Education, 50*(5), 718–731. https://doi.org/10.1080/02602938.2025.2459641

Cui, Y., Schunn, C. D., & Gai, X. (2022). Peer feedback and teacher feedback: A comparative study of revision effectiveness in writing instruction for EFL learners. *Higher Education Research and Development, 41*(6), 1838–1854. https://doi.org/10.1080/07294360.2021.1969541

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers and Education, 210*, Article 104967. https://doi.org/10.1016/j.compedu.2023.104967

Debets, T., Banihashem, S. K., Joosten-Ten Brinke, D., Vos, T. E., de Buy Wenniger, G. M., & Camp, G. (2025). Chatbots in education: A systematic review of objectives, underlying technology and theory, evaluation criteria, and impacts. *Computers & Education*. , Article 105323. https://doi.org/10.1016/j.compedu.2025.105323

Edwards, J., Nguyen, A., Lämsä, J., Sobocinski, M., Whitehead, R., Dang, B., Roberts, A., & Järvelä, S. (2025). Human-AI collaboration: Designing artificial agents to facilitate socially shared regulation among learners. *British Journal of Educational Technology, 56*(2), 712–733. https://doi.org/10.1111/bjet.13534

Ekanayaka, W. I., & Ellis, R. (2020). Does asking learners to revise add to the effect of written corrective feedback on L2 acquisition? *System (Linköping), 94*, Article 102341. https://doi.org/10.1016/j.system.2020.102341

Er, E., Dimitriadis, Y., & Gašević, D. (2021). A collaborative learning approach to dialogic peer feedback: A theoretical framework. *Assessment & Evaluation in Higher Education, 46*(4), 586–600. https://doi.org/10.1080/02602938.2020.1786497

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Fleckenstein, J., Jansen, T., Meyer, J., Trüb, R., Raubach, E. E., & Keller, S. D. (2024). How am I going? Behavioral engagement mediates the effect of individual feedback on writing performance. *Learning and Instruction, 93*, Article 101977. https://doi.org/10.1016/j.learninstruc.2024.101977

Glüsing, R., Fleckenstein, J., Schmidt, F. T. C., & Möller, J. (2026). LLM feedback for academic writing: Effects on students' performance and engagement. *Contemporary Educational Psychology, 85*, Article 102463. https://doi.org/10.1016/j.cedpsych.2026.102463

Graham, S. (2018). Instructional feedback in writing. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 145–168). Cambridge University Press. https://doi.org/10.1017/9781316832134.009.

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

Hadwin, A., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In *In D. H. Schunk & J. A. Greene (Eds.), handbook of self-regulation of learning and performance (2nd ed.pp.83-106). Routledge.* (pp. 83–106). Routledge.

Handley, K., Price, M., & Millar, J. (2011). Beyond "doing time": Investigating the concept of student engagement with feedback. *Oxford Review of Education, 37*(4), 543–560. https://doi.org/10.1080/03054985.2011.604951

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Heil, J., & Ifenthaler, D. (2026). Effects of teacher, peer and self-feedback on student improvement in online assessment: The role of individuals' presumptions and feedback literacy. *Assessment & Evaluation in Higher Education, 51*(2), 281–300. https://doi.org/10.1080/02602938.2025.2530452

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), Article 42. https://doi.org/10.1007/s10648-025-10017-3

Jacobs, S., Kempf, M., & Kiesler, N. (2025). That's not the feedback I need! - Student engagement with GenAI feedback in the tutor kai. In Q. Cutts, S. Sentance, F. McNeill, & C. Alexandru (Eds.), *Proceedings of the 2025 conference on UK and Ireland computing education research* (pp. 1–7). ACM. https://doi.org/10.1145/3754508.3754512.

Jin, X., Jiang, Q., Xiong, W., Feng, Y., & Zhao, W. (2024). Effects of student engagement in peer feedback on writing performance in higher education. *Interactive Learning Environments, 32*(1), 128–143. https://doi.org/10.1080/10494820.2022.2081209

Kim, S., & McCrudden, M. T. (2026). Evaluating the consistency between human raters and three AI systems on the scoring of argumentative essays. *Assessing Writing, 68*, Article 101031. https://doi.org/10.1016/j.asw.2026.101031

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

Lira Lorca, A., Peltzer, K., Busse, V., & Krause, U.-M. (2026). Effects of peer feedback and self-assessment on writing. *Learning and Instruction, 102*, Article 102253. https://doi.org/10.1016/j.learninstruc.2025.102253

Little, T., Dawson, P., Boud, D., & Tai, J. (2025). What does it take to provide effective peer feedback? *Assessment & Evaluation in Higher Education, 50*(5), 775–786. https://doi.org/10.1080/02602938.2025.2475059

Liu, C., & Liang, M. (2026). The effects of AI-assisted feedback on students' perceptions, emotions, and learning outcomes in L2 writing: A three-level meta-analysis. *System (Linköping), 139*, Article 104018. https://doi.org/10.1016/j.system.2026.104018

Lu, Q., Yao, Y., Xiao, L., Zhu, X., & Yin, H. (2026). Exploring the impact of a GenAI-supported feedback practice on student feedback literacy in L2 writing. *Computer Assisted Language Learning*, 1–30. https://doi.org/10.1080/09588221.2025.2605541

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Noroozi, O., Haddadian, G., Gao, X., Schunn, C., Alqassab, M., & Banihashem, S. K. (2025). The value of GenAI for peer feedback provision: Student perceptions and impacts. *International Journal of Educational Technology in Higher Education, 22*(1), Article 61. https://doi.org/10.1186/s41239-025-00558-6

Pack, A., Barrett, A., & Escalante, J. (2024). Large language models and automated essay scoring of English language learner writing: Insights into validity and reliability. *Computers and Education: Artificial Intelligence, 6*, Article 100234. https://doi.org/10.1016/j.caeai.2024.100234

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., & Järvelä, S. (2015). Socially shared regulation of learning: A review. *European Psychologist, 20*(3), 190–203. https://doi.org/10.1027/1016-9040/a000226

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing, 8*(3), 265–289. https://doi.org/10.1016/S1060-3743(99)80117-9

Peltzer, K., Lira Lorca, A., Krause, U.-M., & Busse, V. (2024). Effects of formative feedback on argumentative writing in English and cross-linguistic transfer to German. *Learning and Instruction, 92*, Article 101935. https://doi.org/10.1016/j.learninstruc.2024.101935

Sailer, M., Maier, R., Berger, S., Kastorff, T., & Stegmann, K. (2024). Learning activities in technology-enhanced learning: A systematic review of meta-analyses and second-order meta-analysis in higher education. *Learning and Individual Differences, 112*, Article 102446. https://doi.org/10.1016/j.lindif.2024.102446

Scherer, S., Graham, S., & Busse, V. (2024). How effective is feedback for L1, L2, and FL learners' writing? A meta-analysis. *Learning and Instruction, 93*, Article 101961. https://doi.org/10.1016/j.learninstruc.2024.101961

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction, 91*, Article 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education, 76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Tam, A. C. F. (2025). Interacting with ChatGPT for internal feedback and factors affecting feedback quality. *Assessment & Evaluation in Higher Education, 50*(2), 219–235. https://doi.org/10.1080/02602938.2024.2374485

To, J. (2022). Using learner-centred feedback design to promote students' engagement with feedback. *Higher Education Research and Development, 41*(4), 1309–1324. https://doi.org/10.1080/07294360.2021.1882403

Topping, K. J., Gehringer, E., Khosravi, H., Gudipati, S., Jadhav, K., & Susarla, S. (2025). Enhancing peer assessment with artificial intelligence. *International Journal of Educational Technology in Higher Education, 22*(1). https://doi.org/10.1186/s41239-024-00501-1, 3.

Winstone, N. E., & Carless, D. (2020). *Designing effective feedback processes in higher education : A learning-focused approach.* Routledge.

Winstone, N. E., Gravett, K., Bearman, M., Noble, C., Jensen, L. X., Jones, A., & Nicola-Richmond, K. (2026). The care-full craft of feedback in an age of generative AI. *Assessment & Evaluation in Higher Education*, 1–17. https://doi.org/10.1080/02602938.2026.2643333

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist, 52*(1), 17–37. https://doi.org/10.1080/00461520.2016.1207538

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). The effects of providing and receiving peer feedback on writing performance and learning of secondary school students. *American Educational Research Journal, 58*(3), 492–526. https://doi.org/10.3102/0002831220945266

Wu, Y., & Schunn, C. D. (2023). Passive, active, and constructive engagement with peer feedback: A revised model of learning from peer feedback. *Contemporary Educational Psychology, 73*, Article 102160. https://doi.org/10.1016/j.cedpsych.2023.102160

Yan, D.(A.), Tian, F., Li, H., Gao, Y., & Li, J. (2026). A GenAI-supported dialogic model to enhance the functions of peer feedback in EFL writing: A mixed-method interventional study. *System (Linköping), 136*, Article 103895. https://doi.org/10.1016/j.system.2025.103895

Yu, Q., & Schunn, C. D. (2025). Which experiences are consistently associated with other-regulation of peer feedback length? *Journal of Educational Psychology, 117*(6), 918–938. https://doi.org/10.1037/edu0000913

Zhan, Y., Boud, D., Dawson, P., & Yan, Z. (2025). Generative artificial intelligence as an enabler of student feedback engagement: A framework. *Higher Education Research and Development, 44*(5), 1289–1304. https://doi.org/10.1080/07294360.2025.2476513

Zhan, Y., & Yan, Z. (2026). Students' engagement with ChatGPT feedback: Implications for student feedback literacy in the context of generative artificial intelligence.

Assessment & Evaluation in Higher Education, 51 (5), 821–834. https://doi.org/10.1080/02602938.2025.2471821

Zhang, F., Schunn, C., Chen, S., Li, W., & Li, R. (2023). EFL student engagement with giving peer feedback in academic writing: A longitudinal study. *Journal of English for Academic Purposes, 64*, Article 101255. https://doi.org/10.1016/j.jeap.2023.101255

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13–39). Academic Press.

Zou, D., Xie, H., & Wang, F. L. (2023). Effects of technology enhanced peer, teacher and self-feedback on students' collaborative writing, critical thinking tendency and engagement in learning. *Journal of Computing in Higher Education, 35*(1), 166–185. https://doi.org/10.1007/s12528-022-09337-y