Metacognitive instruction to enhance EFL writers' metacognition, feedback uptake, and revision performance in a GenAI-assisted feedback context

This is the post print version of the article, which has been published in final form at https://doi.org/10.1111/ijal.70328. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

Da (Alex) Yan[1, 2, *]
[1]School of Foreign Language Studies, Wenzhou Medical University, Wenzhou, China.
[2]Wenzhou International Communication Research Center, Wenzhou, China.

Da (Alex) Yan https://orcid.org/0000-0002-1265-9772
* Please address all correspondence to Da (Alex) Yan: alexyan1987@outlook.com

Abstract

This study investigated the impact of a metacognitive instruction (MI) intervention on EFL writers' metacognition, feedback uptake, and revision performance in a Generative AI (GenAI)-assisted feedback environment. The MI intervention integrated prompt engineering into EFL writing pedagogy following a four-step MI sequence. Participants were 141 EFL learners recruited from a Chinese university, with 71 allocated to the experimental group (EG), which received the MI intervention, and 70 assigned to the control group (CG), which engaged with GenAI feedback without additional scaffolding. Following a pretest–posttest quasi-experimental design, data were collected from multiple sources, including questionnaires, reflective journals, think-aloud protocols, and draft and revised writing products. Results showed that the MI intervention significantly improved students' strategy knowledge and all dimensions of metacognitive strategy use. EG participants demonstrated more elaborate uptake of GenAI feedback on global writing issues. Moreover, the EG achieved greater feedback-informed revision gains in content and organization, with no significant differences observed in grammar, vocabulary, and mechanics. This study has important

pedagogical implications for designing and implementing situated metacognitive scaffolding to support EFL writers' critical and effective engagement with GenAI-assisted feedback.

Keywords: EFL writing; AI-assisted feedback; metacognitive instruction; feedback regulation; prompt engineering

1. Introduction

Generative AI (GenAI) for feedback has gained significant attention and implementation within English as a foreign language (EFL) writing classroom (Crosthwaite & Sun, 2026; Li et al., 2024). To a certain extent, GenAI-mediated feedback has served to mitigate teacher workloads amidst the escalating pedagogical demands for the scalable, individualized, and targeted feedback, which is characteristic of contemporary writing instruction (Li, 2025; Zou et al., 2025). However, recent empirical findings caution that the voluminous feedback generated by GenAI, while broadly comparable in quality to human-provided feedback, does not consistently translate into sustained learning gains or engagement (e.g., Deygers et al., 2025; Zou et al., 2025). Noticeably, analyses of students' use of GenAI feedback reveal a tendency toward overly direct application without sufficient modification and adaptation, in contrast to students' responses to traditional teacher feedback (Yang, Shao, et al., 2025; Zhang et al., 2025). Rather than attributing these shortcomings solely to the technology itself, students' insufficient higher-order competencies to critically navigate and leverage AI-generated input may be the primary underlying cause (Nie et al., 2025;

Vieriu & Petrea, 2025).

To address the above deficiency, theoretical frameworks on promoting learner engagement with GenAI feedback (e.g., Zhan et al., 2025) have highlighted that "to facilitate [self-regulation in the context of GenAI]…teachers should … teach students metacognitive strategies … through modelling" (p. 1300). In traditional EFL writing contexts, metacognitive instruction (MI) has been established as an effective facilitator for fostering students' metacognitive awareness and self-regulatory abilities (Teng, 2025). Nevertheless, both the content and pedagogical approaches of MI require adaptation, as emerging evidence suggests that, beyond the general-purpose capabilities needed for effective learning, learners also need context-specific competencies, such as the ability to communicate and collaborate effectively with AI, in technology-enhanced learning environments (Walter, 2024). One promising approach is to integrate prompt engineering into EFL instruction and feedback practices to foster students' ability to effectively navigate and interact with AI throughout the feedback and writing process (Qian, 2025). This adds a practical dimension to traditional MI by equipping students with the skills needed to regulate and monitor GenAI-assisted feedback processes while fostering their metacognitive development over the longer term (Bai et al., 2026).

Unfortunately, situated efforts to adapt MI for GenAI-assisted environments remain sparse, with existing scholarship largely confined to conceptual and theoretical propositions (e.g., Zhan et al., 2025). To the best of the author's knowledge, empirical studies integrating prompt engineering into EFL writing pedagogy remain scarce (Lee

& Palmer, 2025; Qian, 2025), with the limited existing research failing to integrate the development of AI-related competencies with EFL writing instruction within a coherent pedagogical framework (e.g., Woo et al., 2025). This empirical lacuna restricts our understanding of the effectiveness of MI in technology-mediated EFL writing pedagogy and provides little guidance for designing interventions that strengthen learner agency and engagement in AI-augmented writing. More broadly, the lack of empirical evidence on the effectiveness of such context-specific and innovative MI interventions limits our understanding of how teachers can help students mitigate the potential adverse effects of AI technology on learning. To address this research gap, the present study designed an MI intervention tailored to GenAI-assisted writing environments by integrating prompt engineering instruction into an EFL writing course and evaluated its effects on students' metacognitive development, revision-related behaviors, and revision performance.

2. Literature review

2.1. The post-generation challenges facing GenAI feedback: a metacognitive perspective

Since the public release of ChatGPT in late 2022, GenAI-assisted feedback has been increasingly utilized in EFL writing classrooms for feedback (Crosthwaite & Sun, 2026). However, empirically derived insights also highlight students' overly direct and thus ineffective use of GenAI-generated feedback, particularly when tackling complex and demanding tasks. For instance, Chen et al. (2026) reported that EFL writers rejected a significantly higher proportion of feedback at the content level (i.e., global aspects of writing such as content and organization) than at the form level (i.e., surface-level issues such as grammar). In another study, Zou et al. (2025) found that EFL writers demonstrated significantly less successful uptake of GenAI feedback compared to teacher feedback, particularly regarding language use and content issues. Concurrently, research focusing on students' revisions following GenAI feedback highlights fewer effective revision strategies. For instance, Yang et al.'s (2025) analysis revealed that students tend to utilize GenAI feedback to rectify local linguistic issues (i.e., language use, grammar, and mechanics) but are less successful in applying suggestions to global concerns (i.e., organization and content). Similarly, Farrokhnia et al. (2026) found that even though GenAI feedback quality has significantly improved through prompt engineering optimization, students' subsequent revision improvements remained insignificant. They attributed this phenomenon to the nature of students' engagement with GenAI feedback, rather than feedback quality alone.

As evidenced by the above studies, challenges hindering the effective use of GenAI feedback center on how students respond to and apply feedback during revision. In some applied linguistics studies, such as Ene and Upton's (2014), these two behaviors were collectively conceptualized as feedback uptake. In the present study, however, feedback uptake was operationalized exclusively as students' responses to GenAI feedback, drawing on van de Pol et al.'s (2019) classification of students' use of scaffolding. This operationalization is justified for two reasons. First, in GenAI-assisted formative feedback contexts, including the present study and those of Chen et al. (2026) and Zou et al. (2025), students' responses to feedback and their subsequent application of feedback occur at distinct stages: the former during student–GenAI interaction and the latter during a subsequent self-directed revision phase. Second, GenAI-assisted learning is iterative and interactive, enabling students to reconsider, refine, and apply GenAI-generated suggestions even after initially rejecting them (Zhan et al., 2025). Thus, although Ene and Upton's (2014) definition is suitable for teacher feedback contexts, it may not fully capture the dynamic, stage-based nature of GenAI feedback processing during interaction and revision.

The uptake and revision issues in a GenAI-assisted setting are neither new nor unexpected. In Nicol's (2021) framing of internal feedback, successful uptake of external feedback relies on a process of internalization to be transformed into knowledge gains. Similarly, Bolzer et al. (2015) argued that received feedback must be cognitively processed and understood through "mindful processing" to elicit self-adjustment or revision. Thus, in both Nicol's (2021) and Bolzer et al.'s (2015) frameworks, the transformative link between external and internal feedback relies on students' higher-order skills to evaluate, compare, and re-interpret evaluative information. In the landscape of GenAI-assisted feedback, these persistent issues should be urgently addressed, given GenAI's ability to provide a large quantity of practical feedback. According to Zhan et al. (2025), enabling students' regulation of their engagement with GenAI necessitates the fostering and teaching of “metacognitive strategies for planning, monitoring, and evaluation” (p. 1300). Therefore, the above theoretical and practical advancements make metacognitive theories an appropriate lens through which to understand current challenges related to uptake and revision following GenAI feedback.

Metacognition, according to Flavell (1979), encompasses both knowledge and regulatory dimensions. Metacognitive knowledge involves three interacting variables: person variable, which concerns the learner’s awareness of their own cognitive processes and relevant factors; task variable, which relates to the nature and demands of the learning task; and strategy variable, which involves understanding which strategies are effective for achieving specific cognitive goals (Lee & Mak, 2018). Metacognitive regulation enables the real-time configuration of cognitive resources through three core strategies: planning, involving the appropriate selection of strategies and correct allocation of resources before task execution; monitoring, the identification of task comprehension and performance; and evaluation, the appraisal of the regulatory process, the product, and task-performance-related efficiency (Teng, 2020).

In EFL writing, the critical role of metacognition in supporting students as they navigate diverse stages of writing and feedback has been well-established (Lee & Mak, 2018; Teng & Ma, 2024). Notably, EFL writers’ prior experiences naturally foster broader metacognitive skills that can be transferred across domains (Li et al., 2026). However, in emergent and complex scenarios, situated and more explicit metacognitive support may be required (Li et al., 2026; Zhan et al., 2025). Following this reasoning, the observed uptake and revision challenges suggest that EFL writers' current metacognitive repertoires fall short of handling the abundance of GenAI feedback. Thus, we posit that the "post-generation" challenges in GenAI-assisted feedback manifests a lack of specific or situated metacognitive knowledge and capabilities, and we view the proposed solutions as a situated call for critical metacognitive support in EFL writing.

2.2. Metacognitive instruction in EFL writing

A possible approach to enhance students' metacognition during the writing process is MI, which has the potential to foster higher-order cognitive abilities, thereby promoting learning processes and academic performance (Sato, 2023). In EFL writing, extensive research has been conducted employing MI as a pedagogical intervention. Depending on the scaffolding measures utilized (i.e., direct vs. indirect, see Mei et al., 2025), MI interventions have generally followed two approaches.

First, indirect instructional measures, or the integration of metacognitive scaffolds into the learning environment, have been found productive in bolstering metacognitive strategy use. For example, Teng's (2022, 2020) investigations into the effects of metacognitive prompts demonstrated that the combination of metacognitive scaffolds and collaborative writing yielded gains in both metacognitive awareness and writing outcomes. Similarly, a standards-based checklist for EFL writers designed by Mei et al. (2025) was found effective for both perceived and actual metacognitive strategy use, in addition to producing significant increases in writing quality and accuracy. However, researchers also found that such indirect instruction was less effective in promoting metacognitive knowledge compared to the more visible development seen in metacognitive strategy use (Li et al., 2026; Teng, 2020).

Second, another approach to metacognitive scaffolding in EFL writing involves teachers' direct and explicit explanation of metacognitive knowledge and strategies to students and training them to apply them. A representative example is the genre-based instructional assistance in Kessler's (2021) case study, which found that MI scaffolding promoted the rapid development of declarative and procedural genre knowledge through tasks such as rhetorical analysis. However, the corresponding development of conditional knowledge required more practice beyond relatively short-term instructional assistance. In another study, Alfaifi (2022) proposed a five-step instructional model of MI for EFL writing that integrated knowledge orientation, explicit teacher modeling, and student practice. Following its enactment in a Saudi Arabian EFL program, the MI model was found effective in enhancing writing performance and was generally well-received by students. Similar results were found in Wang et al.'s (2025) experimental study, which testified that a training package focusing on the metacognitive knowledge required for collaborative writing resulted in significant improvements across major linguistic dimensions of writing performance.

Collectively, these studies suggest that existing MI interventions tend to emphasize either metacognitive knowledge or metacognitive regulation, but seldom their integration. Such an imbalance may be sufficient for conventional writing tasks but becomes problematic in GenAI-assisted writing, where effective feedback use depends on coordinating what learners know with how they regulate their interactions with AI (Zhan et al., 2025). Accordingly, three limitations constrain the applicability of existing MI to emerging GenAI-assisted writing contexts. First, MI interventions enacted in EFL writing classrooms have often been overly general, focusing on metacognition throughout the entire writing process. This is understandable, as writing itself is "applied metacognition" (Hacker et al., 2009, p. 154). However, it should not be neglected that many subprocesses during writing, for example, the use of GenAI feedback, are also metacognitively intensive activities. In other domains of EFL studies, a trend has emerged to specify the content and focus of MI for specific tasks, particularly classroom interactions. For instance, Sippel (2024) discovered that MI, when combining form-focused and interaction-focused instruction and peer feedback training, was effective in eliciting error-correcting feedback. Similarly, Dao (2020) found that an MI intervention focusing on interaction strategy instruction elevated learner engagement in peer interaction. These studies offer valuable insights applicable to the current context, where engagement and the productive utilization of human-AI feedback interaction are expected.

Second, further efforts should be made to enhance the structure and design of MI within EFL writing. Bozorgian et al. (2025) articulated that incorporating MI with dialogic interaction between learners enabled students to achieve greater performance gains in listening comprehension with reduced cognitive load. Alfaifi's (2022) five-step MI procedure also featured a collaborative practice session before individual practice, which helped students internalize the metacognitive knowledge and skills modeled by teachers. In many previous EFL writing studies, MI has been provided through teacher-fronted sessions, leaving students as passive recipients rather than active participants engaged in the collaborative sense-making necessary for metacognitive development (e.g., Wang et al., 2025, where the MI training package lacked student-centered practice).

Finally, from a methodological perspective, existing empirical studies on MI in EFL writing contexts should embrace a broader range of data collection sources to offer more nuanced and comprehensive understandings. In evaluating writing performance, a "product-oriented" approach has been widely applied. For instance, researchers frequently compared the overall scores of final writing products without multi-dimensional analysis (e.g., Teng, 2020) or focused on specific writing features like CAF metrics (e.g., Mei et al., 2025). Consequently, a "process-oriented" (i.e., focusing on students' behaviors and engagement within the learning/feedback process, such as feedback uptake, revision behavior, and performance gains following feedback) approach should be employed to inform not only performance scores before and after the intervention, but also how learners regulate their engagement with GenAI feedback and how such regulation contributes to subsequent revision outcomes.

2.3. Prompt engineering in GenAI-assisted EFL writing

Prompt engineering, the design of specific instructions to elicit desired outcomes from GenAI, serves as a vital gateway linking learner agency to intended AI performance (Lee & Palmer, 2025). In the broader domain of higher education, a growing number of studies have focused on the significance of prompt engineering in GenAI-assisted learning environments, viewing it as both an emergent critical skill (Qian, 2025) and an important measure to keep GenAI applications aligned with predetermined educational goals (Lee & Palmer, 2025). Beyond improving AI outputs, prompt engineering also requires learners to clarify task goals, monitor AI responses, and iteratively refine their requests, suggesting that it involves substantial metacognitive regulation (Qian, 2025).

However, in EFL research, the integration of prompt engineering has largely been limited to the teacher or designer side. In most cases, such as Zou et al. (2025) and Guo and Wang (2024), students received prompts crafted by their teachers, while neither the prompt design process nor its specific details (e.g., the underlying prompting techniques used) were disclosed. Even in studies where specific prompt design details were discussed, students were neither involved in the prompt engineering process nor educated on how to make better, individualized use of these techniques. For example, while Farrokhnia et al. (2026) applied different prompting techniques (such as zero-shot and chain-of-thought) to generate writing feedback, the students remained mere recipients, lacking access to the rationale and procedures needed to utilize these techniques independently. Although such teacher-designed prompting improves the consistency and quality of AI output, it implicitly positions prompt engineering as a design feature rather than a learner capability, thereby limiting students' agency in regulating GenAI-supported learning (Bai et al., 2026).

To the best of our knowledge, the only EFL writing study that explicitly taught prompt engineering techniques was conducted by Woo et al. (2025). Their study, which exposed 27 Hong Kong tertiary students to a 100-minute workshop on prompt engineering, demonstrated that this targeted instructional intervention significantly enhanced students' self-efficacy, understanding of AI concepts, and prompt engineering skills. Nevertheless, the instructional focus remained on AI literacy and prompting competence, leaving unanswered whether these gains translated into more effective engagement with GenAI-generated feedback or improved writing revision.

Therefore, the current study designs an innovative MI intervention. Rather than treating prompt engineering as an isolated technical skill, we conceptualize it as a situated form of metacognitive regulation through which learners plan information needs, monitor AI responses, evaluate feedback quality, and refine subsequent interactions. The rationale for this MI design was informed by evidence suggesting that prompt engineering is a process-oriented practice that not only aligns with the development of learners' regulatory and reflective capabilities, but also serves as an effective scaffolding strategy for fostering metacognitive development in technology-enhanced learning environments (Bai et al., 2026; Qian, 2025). Drawing on established design models and principles (e.g., Alfaifi, 2022; Bozorgian et al., 2025), we developed an MI framework that interconnects these insights within a pedagogical loop of “teacher modeling - collaborative sense-making - individual reflection.”

2.4. Research questions

To address the identified research gaps, this study examined the effects of the MI intervention on students’ metacognition, feedback uptake, and revision performance in GenAI-assisted writing. We operationalized metacognition not only as a conceptual compound of metacognitive knowledge and strategy use (Teng, 2020), but also as a combination of knowledge and capabilities in both self-perception and actual use during the writing process (Mei et al., 2025). Furthermore, adopting a process-oriented approach to writing, we intend to closely examine students’ uptake of GenAI feedback and gains attained through revision across different writing indices (e.g., content, organization, and grammar). The rationale for this focus lies in the above-reviewed “post-generation challenges” facing EFL writers, alongside criticisms that GenAI usage often fails to elicit significant learning gains, particularly regarding higher-order writing dimensions. Consequently, this study aims to design and implement an intervention that integrates the instruction of prompt engineering skills within the established framework of MI in an EFL writing context, while examining students’ performance and behaviors across these key observable variables under both treatment and non-treatment conditions. To achieve this, the following research questions will be addressed:

RQ1: To what extent does the MI intervention affect students’ metacognitive knowledge and strategy use?

RQ2: To what extent does the MI intervention affect students' feedback uptake?

RQ3: To what extent does the MI intervention affect students' revision performance?

3. Methodology

3.1. Design

The current study was part of a larger research project on GenAI-assisted feedback in an EFL writing context. This quasi-experimental study employed a pretest–posttest design. Specifically, we adopted a multi-source approach to data collection: (1) to examine students' metacognition, we administered a questionnaire and collected reflective journals; (2) to investigate feedback uptake, we analyzed students' think-aloud protocols of their interactions with GenAI feedback; and (3) to evaluate revision performance, we examined feedback-mediated performance gains based on students' drafts and revised essays. The reasons for adopting this approach are threefold: (1) the integration of both qualitative and quantitative analysis methods increases the opportunity to obtain an in-depth and comprehensive interpretation of the results (Creswell & Clark, 2006); (2) previous studies have warned against the exclusive use of metacognitive questionnaires to avoid the excessive bias inherent in self-reported perceptual data (Li et al., 2026; Mei et al., 2025); and (3) this approach aligns with the "process-oriented" framework for evaluating writing performance argued for in previous sections.

3.2. Participant

The participant pool consisted of 183 second-year undergraduate students (from four classes) enrolled in a four-year program of educational science at a Chinese university. To ensure the study had sufficient statistical power to detect differences in these practical outcomes, an a priori power analysis was conducted using G*Power software. For a one-way analysis of covariance (ANCOVA) with two groups and one covariate, the calculation based on a medium effect size (f = 0.25), a significance level of α = 0.05, and a desired statistical power of 0.80, indicated a minimum required sample size of 128 participants.

With the initial sample pool sufficiently large to buffer against potential attrition and provide adequate statistical power, we assigned Classes 1 and 3 to the control group (CG; n = 91) and Classes 2 and 4 to the experimental group (EG; n = 92). During the study, 14 students from the CG and 13 from the EG withdrew due to personal or medical reasons. Additionally, 7 students from the CG and 8 from the EG failed to provide complete data as requested. This resulted in a final sample of 141 participants (CG: 70; EG: 71). No significant differences were found between groups regarding gender [N_{female_cg}=59; N_{female_eg}=57], age [M_{age_cg}=19.4 years; M_{age_eg}=19.6 years], or writing proficiency[1] [$F(1, 139) = 0.83$, p= 0.364]. Furthermore, pre-study survey (5-point scale)

[1] Baseline writing proficiency was assessed prior to the semester using an official College English Test Band 4 (CET-4), a national examination for college students in China.

indicated that both groups had similar experiences with formative feedback [M=3.5, SD=1.1; $F(1, 139) = 0.13$, $p = .72$] and reported a low level of familiarity with AI feedback [M=1.72, SD=0.8; $F(1,139)=0.23$, $p=.63$].

3.3. GenAI configuration

The large language model (LLM) used in this research was Alibaba's Qwen3-4B. This model was selected for several reasons. First, Qwen provides multilingual language processing capabilities well suited to EFL writing contexts (Yang, Li, et al., 2025). Second, it has been adopted in previous studies of GenAI-supported EFL writing and feedback by the research team in this instructional setting. Moreover, its accessibility and relatively affordable API access make it suitable for sustained classroom implementation and promote methodological consistency across studies. Third, the LLM served as a consistent technological platform to support students' metacognitive regulation of feedback processing and revision. Therefore, model-specific differences in performance were not a primary concern in this study.

Students in both groups accessed the LLM via Cherry Studio, a popular GenAI client offering greater configuration flexibility than the official web-based interface. The AI model was configured with a low temperature setting (0.3), which controls the variability of its responses. Lower temperature values produce more consistent outputs and reduce variation across similar inputs. Following the recommendations of Coskun and Er (2026), this setting was adopted to ensure consistent feedback quality throughout the study.

3.4. The MI intervention

The sequences within the MI intervention are shown in **Table 1**.

Table 1. Content of the MI intervention.

Phase	**Activities**	**Materials & Scaffolds**	**Purpose**
Step 1: Orientation (20 minutes)	• Defining metacognitive strategies • Explaining the utility of prompt engineering in AI-assisted writing	Introductory explanations	Building metacognitive knowledge
Step 2: Teacher Modelling (60 minutes)	• Explicit demonstration of strategies as learnable actions • Think-aloud approach articulating the details and affordances of each strategy	•3 video recordings of human–AI feedback encounters via a MOOC platform (using 3 distinct writing samples) •Prompt engineering technique list from Woo et al. (2025).	Fostering strategy use mastery
Step 3: Collaborative Practice (Two 40-min sessions)	• Scaffolded practice under teacher supervision • Engaging with AI to seek feedback using the writing samples from Step 2	• Step 2 writing samples • “Think sheets”	Collaborative sense-making:
Step 4: Individual Practice & Reflection (20 minutes)	• Independent AI-assisted writing practice • Writing a reflection note summarizing gains and strategy development	• A new writing sample • Step 3 instructional materials & “think sheets”	Autonomous application

As illustrated, the four-step procedure was adapted from existing MI models (e.g., Alfaifi, 2022; Bozorgian et al., 2025) to encourage autonomous strategy internalization. The most critical component of this MI intervention is teacher modelling, where a detailed explanation of metacognitive knowledge, prompt engineering skills, and feedback regulation strategies is provided through demonstration (see **Appendix S1** for the modelling plan and script excerpts). Specifically, this session focuses on aligning instructions on metacognitive knowledge, prompt engineering techniques, and feedback regulation within a unified instructional framework. Consequently, the instruction of technology-based techniques is integrated with, rather than isolated from, the teaching of writing and evaluation skills for EFL writers.

3.5. Measures

3.5.1 Pre- and post-intervention writing tests

The study utilized two comparable writing tasks selected from CET-4 test banks. The topics of the two prompts were "how to handle the teacher-student relationship" and "how to maintain the doctor-patient relationship." In accordance with official guidelines, students were required to complete an essay of at least 120 words within 30 minutes. Notably, we did not employ the official CET-4 grading rubric for evaluation, as its holistic impression-based scheme risked excessive subjectivity and provided limited analysis across specific linguistic dimensions. Instead, we used a 100-point analytic rating instrument devised in a prior study (Author, 2024), which assesses five distinct dimensions: content, organization, grammar, vocabulary, and mechanics.

Given our focus on feedback-mediated revision performance, both the initial drafts and revised versions of each essay were collected and scored by two experienced EFL writing teachers. The grading demonstrated a high level of inter-rater reliability [r = 0.86, $p < .001$]. Revision performance was calculated by subtracting the draft scores from the revised essay scores. To ensure these gains validly represented improvements specifically informed by GenAI feedback, revisions were conducted on Shimo.im, an online collaborative office suite, with the “track changes” feature enabled. Raters then manually processed the revised versions by filtering out any edits that were not directly grounded in the GenAI feedback comments.

3.5.2 Metacognitive strategies in writing

The Metacognitive Strategies in Writing (MSW) questionnaire, developed and validated by Qin et al. (2022), was adopted to assess students’ development of metacognitive knowledge and strategy use in EFL writing. The MSW scale consists of 30 items measured on a 7-point Likert scale (1 = strongly disagree; 7 = strongly agree). The scale includes six dimensions: person, task, strategy, planning, monitoring, and evaluating. The first three dimensions (person, task, and strategy) are subsumed under metacognitive knowledge, while the remaining three (planning, monitoring, and evaluating) are categorized under metacognitive strategy use. The reliability of the questionnaire was assessed using Cronbach’s alpha [.82 and .86 for pre- and post-test], on par with the reliability coefficients reported by the original researchers (Qin et al., 2022).

3.5.3 Reflective journaling

Reflective journaling, sometimes referred to as a written account, has been an effective approach for tracking students' thought processes in previous research (e.g., Mei et al., 2025; Teng, 2020). We collected students' reflective journals after the post-test writing task to provide supplementary qualitative insights into the effects of the MI intervention on metacognition. Students were required to respond to prompts designed to elicit reflection on whether and which specific metacognitive knowledge and strategies were used at different stages of their interaction with GenAI.

3.5.4 Think-aloud protocol

We invited all students to participate in an audio-recorded think-aloud session following the post-test, focusing on their explanations of GenAI feedback uptake. Prior to the study, students were trained in the think-aloud protocol and instructed to provide detailed, comprehensive verbalizations. Following data cleaning and audio trimming, 420 minutes of audio data were transcribed for coding. A coding scheme informed by van de Pol et al. (2019) was adopted, encompassing three categories of feedback uptake: (1) ignore, (2) elaboration/modification, and (3) apply. The coding process involved extracting focal explanations from student verbalizations, forming codes based on key insights, and categorization. Three EFL teachers coded the data, achieving high inter-rater reliability (r = 0.92).

3.6. Procedures

The research was embedded within a non-English major writing course (**Figure 1**). During week 1, students in both groups received an orientation that included a brief introduction to the course's curricular and pedagogical details, tailored training aimed at facilitating student use of GenAI for writing feedback, and an introduction to the rationale and procedures of the study. Regarding writing instruction, the course featured a theme-based curriculum, which included four weeks of instruction on general writing strategies (e.g., diction, phrasing, sentences, and paragraphs), followed by dedicated writing and feedback weeks (i.e., weeks 6 and 11). During these writing and feedback weeks, students drafted their essays, sought feedback from GenAI, and applied that feedback to their subsequent revisions. To maintain consistency, the curricular settings and pedagogical environments were identical for both groups, and both were taught by the same instructor. MI for the EG was offered as extracurricular activities so that classroom teaching sessions were not compromised.

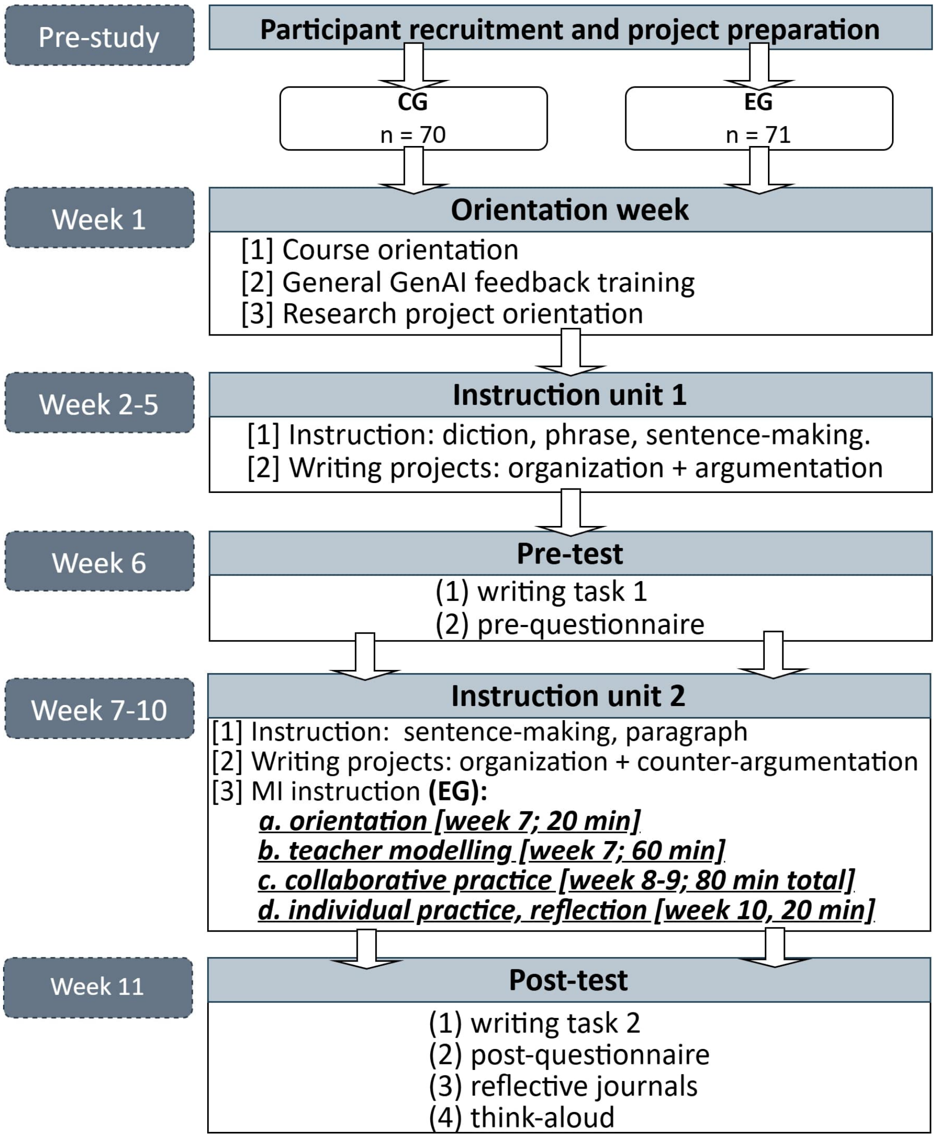

Figure 1. Research procedures.

During week 6, students' drafts and revised essays were submitted to the instructor for grading. In addition, a pre-test questionnaire on metacognitive strategies was administered. Correspondingly, similar post-test data (including the draft and revised essays for task 2, alongside the post-test questionnaire) were collected in week 11. Moreover, students' reflective journals and think-aloud verbalizations were collected following the completion of task 2 in week 11.

3.7. Analysis

To determine the effects of the MI intervention on revision and metacognition, a one-way ANCOVA was performed using R. This model allowed for a robust comparison of post-intervention outcomes after controlling for baseline differences. Prior to the main analysis, we verified the underlying assumptions of ANCOVA tests. Specifically, Levene's test was used to evaluate the homogeneity of variance, yielding non-significant results for both the questionnaire data (p = .38) and revision performance (p = .17). Additionally, the assumption of homogeneity of regression slopes was satisfied, as evidenced by non-significant covariate × group interactions [ps>.05]. To compare the proportions of various feedback uptake conditions between the two groups during the post-test essay revisions, categorized results were calculated and presented. A chi-square test was subsequently performed to examine the effects of the intervention on the distribution of uptake strategies between the two groups.

Qualitative data from journaling were thematically analyzed following the framework established by Braun and Clarke (2006). To ensure the credibility of the findings, trustworthiness was maintained through member-checking (Birt et al., 2016). Following a systematic process of iterative theme identification and code refinement, the resulting interpretive narratives and excerpts were triangulated with the quantitative results.

4. Results

4.1. Quantitative analyses

4.1.1 Analysis of questionnaires data

Descriptive statistics for the questionnaire data are provided in **Table 2**. Results of the one-way ANCOVA revealed that, after controlling for pre-survey scores, there were statistically significant differences in planning [$F(1,138)=356.2$; $p<.001$; $\eta^2_p=0.72$], monitoring [$F(1,138)=909.1$; $p<.001$; $\eta^2_p=0.87$] and evaluating [$F(1,138)=94.6$; $p<.001$; $\eta^2_p=0.41$] strategy use, as well as strategy knowledge [$F(1,138)=723.4$; $p<.001$; $\eta^2_p=0.84$]. However, significant differences were not found in person [$F(1,138)=1.6$; $p=.2$; $\eta^2_p=0.01$] and task [$F(1,138)=2.6$; $p=.11$; $\eta^2_p=0.02$] knowledge measures.

Table 2. Descriptive statistics of questionnaire data.

Measures	Pre-survey M(SD)		Post-survey M(SD)	
	CG	EG	CG	EG
Person	2.2 (0.7)	2.5 (0.8)	2.5 (0.5)	2.8 (0.9)
Task	4.5 (1.1)	4.0 (1.3)	5.1 (1.4)	4.6 (1.3)
Strategy	3.6 (1.2)	3.9 (0.9)	4.3 (1.0)	5.7 (1.1)
Planning	4.3 (0.9)	4.4 (1.1)	4.7 (1.1)	5.2 (1.6)
Monitoring	3.6 (0.8)	3.7 (0.9)	4.1 (1.1)	5.0 (1.2)
Evaluating	4.0 (0.9)	4.4 (1.1)	4.5 (1.2)	5.2 (1.3)

4.1.2 Analysis of feedback uptake

The distribution of feedback uptake (see **Appendix S2** for samples) across the five writing dimensions is shown in **Table 3**. Based on the results from the chi-square tests, statistically significant associations were observed for the dimensions of content [$\chi^2(2, N = 253) = 32.15$, $p < .001$, $V = .35$], and organization [$\chi^2(2, N = 263) = 59.91$, $p < .001$,

V = .48]. Descriptive statistics revealed that while the CG predominantly "applied" feedback (47.3% for content; 63% for organization), the EG showed a significantly higher tendency to "elaborate/modify" their work (62.9% for content; 66.7% for organization). In contrast, no significant differences were found for dimensions including grammar [χ^2 (2, N = 250) = 5.55, p = .06, V = .15], vocabulary [χ^2(2, N = 280) = 0.95, p = .62, V = .06], and mechanics [χ^2 (2, N = 379) = 3.77, p = .15, V = .10]. Across these technical categories, both groups maintained high "apply" rates, particularly in mechanics where the CG (66.4%) and EG (74.4%) both favored direct application over modification. These results suggest that the MI intervention successfully fostered deeper revision strategies for global writing concerns but had less significant impact on local, surface-level corrections.

Table 3. Proportional distribution of feedback uptake and chi-square test results.

Writing dimensions	CG: Uptake n (%)			EG: Uptake n (%)			Chi-square test
	A	E/M	I	A	E/M	I	
Content	61 (47.3%)	36 (27.9%)	32 (24.8%)	34 (27.4%)	78 (62.9%)	12 (9.7%)	χ^2= 32.15 p < .001 V = .35
Organization	75 (63%)	23 (19.3%)	21 (17.7%)	41 (28.5%)	96 (66.7%)	7 (4.8%)	χ^2= 59.91 p < .001 V = .48
Grammar	52 (43%)	46 (38%)	23 (19%)	49 (37.1%)	68 (51.5%)	15 (11.4%)	χ^2= 5.55 p= .06 V = .15
Vocabulary	68 (46.3%)	59 (40.1%)	20 (13.6%)	55 (41.4%)	61 (45.9%)	17 (12.7%)	χ^2= 0.95 p = .62 V = .06
Mechanics	140 (66.4%)	33 (15.6%)	38 (18%)	125 (74.4%)	24 (14.3%)	19 (11.3%)	χ^2= 3.77 p = .15 V = .1

Note: A: apply; E/M: elaboration/modification; I: ignore; V: Cramer's V.

4.1.3 Analysis of revision performance

Descriptive statistics for writing revision performance after revision are shown in **Table 4**. Results of the one-way ANCOVA revealed that, after controlling for pre-test revision performance scores, there were statistically significant differences between the two groups in content [$F(1,138)= 13091.7$; $p<.001$; $\eta^2_p=0.99$] and organization [$F(1,138)= 3339.4$; $p<.001$; $\eta^2_p=0.96$] measures. However, for grammar [$F(1,138)=0.4$; $p=.52$; $\eta^2_p=0.003$], vocabulary [$F(1,138)=2.6$; $p=.11$; $\eta^2_p=0.02$], and mechanics [$F(1,138)= 0.5$; $p=.5$; $\eta^2_p=0.004$] dimensions of writing, no statistically significant differences were found.

Table 4. Descriptive statistics of revision performance.

Measures	Pre-test M(SD)		Post-test M(SD)	
	CG	EG	CG	EG
Content	3.2 (1.1)	3.8 (1.2)	4.0 (1.3)	7.1 (1.2)
Organization	2.1 (1.0)	3.2 (1.3)	3.3 (1.2)	6.4 (1.1)
Grammar	5.5 (1.2)	6.1 (1.5)	7.8 (1.1)	8.4 (1.6)
Vocabulary	4.1 (1.3)	3.6 (1.1)	4.9 (1.6)	4.6 (1.4)
Mechanics	2.8 (1.5)	3.4 (1.4)	3.7 (1.4)	4.4 (1.2)

4.2. Qualitative analysis of journal data

The following three themes emerged from the journal analysis: (1) making plans; (2) controlling the quality of AI feedback; and (3) navigating the conversations with AI (see **Appendix S3** for the codebook). It should be noted that we did not present the frequency of the codes; instead, inspired by the methodology of Mei et al. (2025), we recorded the number of students (termed "sources") whose behaviors and narratives fell within the corresponding categories.

4.2.1 Making plans

Students in both groups commonly engaged in deliberate forethought. The most frequent behavior of students from the two groups involved **incorporating task-specific information** (code 1.3), where students defined the parameters of the assignment to AI. Beyond the task itself, students frequently engaged in **pre-interaction strategy choice** (1.4). This involved planning how to use the AI as a cognitive tool, with students noting, for instance, *"I will request an outline first to see the structure."* While these task- and strategy-related behaviors were prevalent in both groups, EG students showed a higher tendency to sequence these steps logically compared to the CG.

Comparatively, planning centered on the person variable was significantly lower. In particular, **incorporating personal preferences** (1.2) remained the least utilized code, with only a few students specifying their own limitations, such as: *"Make it quick since I have only 30 minutes to finish the whole writing."* Moreover, more EG members used **setting a persona** (1.1), as illustrated by prompts such as *"Pretend you are an expert medical researcher."* In contrast, CG learners more frequently initiated interactions with GenAI without specifying a particular role for the model to assume.

4.2.2 Controlling quality of AI feedback

While quality control occurred in both groups, the EG demonstrated a much higher frequency of "Evaluating + Monitoring" behaviors. For **accuracy verification** (2.1), EG students actively questioned the AI's validity, noting: *"I checked this against the external materials to see if the AI was hallucinating."* Regarding **relevance and depth judgment** (2.2), EG students were more critical than CG learners. For instance, an EG student stated that *"The response was too superficial for a university level."* This led to a higher rate of **identifying AI limitations** (2.3), where students explicitly documented flaws: *"It used a misleadingly confident tone for a complex topic."* Consequently, **feedback evaluation and iteration** (2.4) became an iterative loop for the EG, characterized by quotes like: *"I told AI to re-evaluate the third paragraph as I made some adjustments to the tone."*

4.2.3 Navigating conversation with AI

The EG excelled in real-time regulation during the "back-and-forth" process. **Conversation monitoring** (3.1) was a hallmark of the EG, with students actively tracking the dialogue's trajectory: *"I had to restate my question because the AI forgot the context."* Once a drift or gap was identified, students engaged in **prompt adaptation** (3.2) to refine the output quality. For example, a student commented that *"The previous rubric in the requirement could be changed as follows so that we can focus on bigger areas in the writing."* This was often paired with **adaptive strategy regulation** (3.3), where students pivoted their approach based on the AI's performance: *"The broad explanation didn't help, so I switched to asking for a real-world example."* Finally, **persistence and decision-making** (3.4) showed how EG students took agency over the interaction. They either stayed committed: *"I kept iterating until I finally* *understood the suggestion"* or knew when to stop: *"The AI was becoming unproductive and repetitive, so I decided to disengage and use the library instead."*

5. Discussion

5.1. Effects on metacognition

The study suggests that the situated MI intervention, which integrated explicit prompt engineering instruction with metacognitive strategy training, facilitated the development across multiple domains of metacognitive knowledge and strategy use. These comprehensive effects align with empirical insights from previous research (Mei et al., 2025; Teng, 2022, 2020). More importantly, the findings extend this literature by suggesting that prompt engineering can function as a practical medium through which learners enact metacognitive regulation during GenAI-assisted writing (Bai et al., 2026; Qian, 2025). Further investigation into students' reflective journals revealed that students who were exposed to the MI intervention more effectively engaged in monitoring, evaluating, regulating, and adjusting their use of GenAI for feedback. This reinforces the notion that metacognitive support is essential for productive interaction with feedback and writing (e.g., Dao, 2020; Sippel, 2024). Thus, this study demonstrates that MI is a direct approach to supporting students' development of metacognitive strategy use in EFL writing (Lee & Mak, 2018; Li et al., 2026), particularly when metacognitive strategies are operationalized through prompt engineering practices that support human–AI interaction, potentially addressing the "superficial" use of GenAI recently noted in the literature (e.g., Yang, Shao, et al., 2025; Zhang et al., 2025).

Regarding metacognitive knowledge, the results for each subordinate dimension offer important insights. First, MI promoted students' mastery of strategy knowledge during their interactions with GenAI for feedback. This development is encouraging, as the literature (e.g., Kessler, 2021) suggests that fostering EFL writers' strategic or conditional knowledge is notoriously difficult and usually requires extended support. EG members not only reported greater mastery of strategy knowledge but also translated this knowledge into purposeful prompt engineering practices. These practices suggest that prompt engineering functioned as an externalized form of metacognitive regulation, enabling learners to plan interactions with GenAI, monitor AI responses, evaluate their usefulness, and adjust subsequent prompts accordingly (Bai et al., 2026). This success likely stems from the MI's design. Content-wise, our operationalization of metacognitive strategies, including teacher modeling of human-AI interactions, provided a practical roadmap for students. In contrast, previous efforts (e.g., Kessler, 2021; Teng, 2020) often lacked such direct, contextualized scaffolding, leading to limited gains. Structurally, the structural configuration of the MI, drawing on lessons from recent studies (e.g., Alfaifi, 2022; Bozorgian et al., 2025), enabled students to internalize strategic knowledge for immediate use, even within a limited timeframe (cf. Kessler, 2021). Second, questionnaire and journal data revealed that while students possessed strong task knowledge, they lacked mastery of the person variable. Both groups demonstrated awareness of incorporating task-specific requirements when prompting GenAI. This echoes Li et al. (2026), suggesting that metacognitive awareness from prior EFL writing and feedback experiences can transfer to new domains, such as goal-setting for human-AI interaction. Unfortunately, students did not systematically integrate their self-perceptions of cognitive capabilities throughout the feedback interactions. This deficiency may stem from their lack of experience with individualized feedback, which has often been unrealistic in teacher-fronted environments due to logistical constraints, and from the lack of person-centric communication (Li et al., 2026).

5.2. Effects on feedback uptake

Analysis of feedback uptake revealed an interesting pattern: while both groups shared a direct approach to feedback comments on local issues, EG learners demonstrated a more sophisticated application of feedback on global issues compared to their CG peers. The feedback uptake behaviors of the CG echoed observations from previous studies that students tend to directly apply corrections suggested by GenAI without sufficient reconsideration (Zhang et al., 2025; Zou et al., 2025).

In contrast, the way EG learners processed GenAI feedback on global issues warrants closer attention. Their responses suggest that feedback uptake in GenAI-assisted environments is more dynamic and cognitively demanding than that in traditional feedback contexts. As the qualitative findings showed (4.2.2), students actively evaluated the correctness and appropriateness of GenAI feedback rather than accepting it uncritically. This indicates that feedback uptake is itself a learning process through which students construct knowledge via critical engagement. Consequently, the construct of feedback uptake, as defined by applied linguists such as Ene and Upton (2014), may need to be revisited. Their definition primarily conceptualizes uptake as learners' revision decisions and outcomes following feedback. By contrast, Zhan et al. (2025) conceptualized GenAI-assisted feedback as a cyclical self-regulated learning process involving goal setting, monitoring, and reflection. From this perspective, uptake should encompass learners' critical evaluation, verification, and iterative refinement of AI-generated feedback, rather than only their subsequent revision decisions. To this extent, our findings support distinguishing learners' responses to feedback from revision itself, as proposed in our operationalization of feedback uptake.

From a pragmatic perspective, students who received metacognition-based prompt engineering scaffolding demonstrated greater awareness of potential AI hallucinations and algorithmic bias, two major sources of inaccurate AI output (Bearman et al., 2024). This finding aligns with evidence that prompt engineering can raise learner awareness of bias in "high-stakes and value-laden scenarios" (Qian, 2025, p. 1792). Previous research has likewise shown that, without metacognitive or technical support, EFL learners often struggle to effectively utilize GenAI feedback on global writing issues such as content (e.g., Zou et al., 2025). The present findings therefore suggest that integrating prompt engineering instruction into metacognitive writing pedagogy can promote more active and critical engagement with GenAI feedback, enabling learning through iterative verification, inquiry, and refinement (Zhan et al., 2025).

5.3. Effects on revision performance

The analysis of students' revision performance indicates that the MI intervention effectively improved overall writing performance. This finding aligns with previous literature regarding the positive relationship between metacognitive support and EFL writing quality (Mei et al., 2025; Teng, 2020). Notably, the improvement in writing performance was accompanied by more substantial feedback-informed revisions of global-level issues. Specifically, the study found that the intervention triggered a more significant increase in the scores of feedback-informed revision performance among EG members in global-level issue corrections. However, regarding the increase resulting from revising local issues, the two groups showed no significant differences. This result was not unexpected, since previous studies, such as Yang et al. (2025), reported that students tend to successfully apply suggestions on local issues regardless of specialized support.

The greater improvement in global-level issue correction suggests that the MI intervention helped students engage more deeply with GenAI feedback and address more complex aspects of writing development. This finding also highlights the continued relevance of explicit metacognitive scaffolding in GenAI-assisted writing environments. Specifically, although GenAI can provide readily accessible feedback, students still require support in processing, evaluating, and appropriately applying such feedback. Thus, the MI intervention introduced here contributes to addressing concerns that the overly straightforward use of GenAI might hinder rather than augment the development of writing proficiency (Deygers et al., 2025; Zhang et al., 2025). More broadly, this also suggests that MI, or teacher-led direct and explicit metacognitive scaffolding, remains highly relevant and effective in an era of rapid technological innovation (Zhan et al., 2025).

Nevertheless, the intervention effects on content and organization revisions warrant cautious interpretation. Although the ANCOVA results revealed extremely large effect sizes for these dimensions, such values should be considered in light of the operationalization and distributional characteristics of revision performance. Because revision performance was calculated as gain scores between revised versions and drafts, restricted variability in these scores may have amplified apparent between-group differences, as standardized effect sizes are influenced by the magnitude of group differences relative to score variability (Lakens, 2013). Supporting this interpretation, additional ANCOVA analyses based on holistic writing scores, rather than gain scores, showed large but reasonable effects (η^2_p = .36 for content and .45 for organization), indicating that the measurement approach partly contributed to the magnitude of the observed effects. Moreover, qualitative and process-based findings corroborate the substantive nature of these gains. Compared with CG students, EG students demonstrated greater uptake of AI feedback on content and organization (see 4.1.2), conducted more extensive revisions (i.e., greater revision length and frequency), and engaged more critically with AI feedback by evaluating its relevance, recognizing its limitations, and refining their interactions with AI (see 4.2.2 and 4.2.3).

5.4. Implications

The findings of this study provide several implications for EFL writing teachers engaging with GenAI-assisted writing. First, AI competence should be integrated with, rather than separated from, EFL writing pedagogy. Specifically, prompt engineering should be viewed not as an isolated technical skill, but part of the overall competence that enables students to regulate their interaction with GenAI. For example, students can be guided to construct prompts for different writing purposes, such as generating ideas during planning, examining argument development during drafting, and obtaining targeted feedback during revision. Importantly, such instruction should emphasize not only how to elicit useful AI responses but also how to critically evaluate and refine them. Through these practices, prompt engineering becomes a writing-oriented strategy that enables students to regulate their engagement with GenAI.

Second, MI-based scaffolding should be adapted to address the emerging demands of GenAI-mediated writing. While traditional MI in EFL writing primarily targets students' regulation of their own writing processes, GenAI-assisted writing requires additional support for regulating interactions with AI, a process that itself involves substantial metacognitive activity. Therefore, MI should explicitly address how students plan AI use, monitor AI responses, and evaluate whether feedback should be applied, modified, or rejected. This is particularly important because the current study found that MI-supported students were more likely to elaborate on and modify AI feedback, in addition to resisting hallucinations and inaccuracies in AI feedback. Teachers may foster such abilities through structured MI sequences involving teacher modelling, collaborative practice, individual application, and reflection. Think-aloud demonstrations of human–AI feedback interactions, for example, can help students observe how experienced writers identify limitations in AI feedback, adjust prompts, and make revision decisions.

Third, attention should be given to student agency in specific subprocesses of EFL writing and feedback activities under GenAI-mediated conditions. Teachers should recognize the importance of students' active regulation during feedback interactions, particularly when determining the reliability and applicability of AI-generated input. For instance, feedback uptake should be conceptualized as an iterative process through which students interpret, evaluate, and transform AI-generated input rather than simply accept or reject suggestions. Writing activities should therefore provide opportunities for students to articulate their feedback-related reasoning, such as explaining why particular AI suggestions were adopted, modified, or rejected, and what evidence informed their decisions. Such practices can help students develop the agency and regulatory capacities needed to critically engage with GenAI feedback, including recognizing potential hallucinations, questioning uncertain outputs, and maintaining responsibility for final revision decisions.

5.5. Limitations and future prospects

First, participants were recruited from a single Chinese university, forming a relatively small and homogeneous sample of EFL learners. Future studies should recruit larger, more diverse samples across multiple institutions and settings to strengthen external validity. Second, although a multi-source data collection approach was adopted, space constraints prevented an in-depth analysis of the micro-level nuances of human–AI interaction and feedback-informed revision. Future research could employ stimulated recall, process-tracing methods, or longitudinal tracking of prompt-engineering behaviors to capture these dynamics more fully. Third, as a pioneering effort, the MI intervention design warrants further refinement. Future research should explore how teacher scaffolding can better support students in evaluating and responding to low-quality AI outputs, including hallucinated or biased information. In particular, more attention should be given to how teacher modelling, guided practice, and reflection can cultivate students' ability to verify AI-generated content and make informed decisions during human–AI interactions.

References

Alfaifi, M. J. (2022). A suggested model for metacognitive strategy instruction in EFL writing classrooms. *Reading & Writing Quarterly*, *38*(4), 323–339. https://doi.org/10.1080/10573569.2021.1954569

Bai, S., Yeung, S. S., & Lo, C. K. (2026). Enhancing the effect of AI-assisted learning: The use of scaffolding strategies to develop students' prompt engineering skills. *Interactive Learning Environments,* 1–22. https://doi.org/10.1080/10494820.2026.2649547

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education,* 1–13. https://doi.org/10.1080/02602938.2024.2335321

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? *Qualitative Health Research*, *26*(13), 1802–1811. https://doi.org/10.1177/1049732316654870

Bolzer, M., Strijbos, J. w., & Fischer, F. (2015). Inferring mindful cognitive-processing of peer-feedback via eye-tracking: Role of feedback-characteristics, fixation-durations and transitions. *Journal of Computer Assisted Learning*, *31*(5), 422–434. https://doi.org/10.1111/jcal.12091

Bozorgian, H., Muhammadpour, M., & Alehchali, A. B. (2025). The effects of metacognitive intervention and metacognitive intervention through dialogic interaction on L2 learners' listening comprehension and cognitive load. *Language Awareness*, 1–27. https://doi.org/10.1080/09658416.2025.2489937

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Coskun, A., & Er, E. (2026). Hybrid intelligence feedback: A model proposal and preliminary study. *Journal of Computer Assisted Learning*, *42*(2), e70212. https://doi.org/10.1002/jcal.70212

Creswell, J. W., & Clark, D. V. L. P. (2006). *Designing and Conducting Mixed Methods Research* (1st edition). Sage Publications, Inc.

Crosthwaite, P., & Sun, S. (2026). Generative AI and L2 Written Feedback Studies: A Scoping Review. *RELC Journal*, *57*(1), 207–219. https://doi.org/10.1177/00336882251386530

Dao, P. (2020). Effect of interaction strategy instruction on learner engagement in peer interaction. *System*, *91*, 102244. https://doi.org/10.1016/j.system.2020.102244

Deygers, B., Buelens, L., Chan, D., Schildt, L., Van Parys, A., & Vanbuel, M. (2025). The impact of automated writing evaluation on writing gains. *ELT Journal*, *79*(3), 352–362. https://doi.org/10.1093/elt/ccaf020

Ene, E., & Upton, T. A. (2014). Learner uptake of teacher electronic feedback in ESL composition. *System*, *46*, 80–95. https://doi.org/10.1016/j.system.2014.07.011

Farrokhnia, M., Latifi, S., Papadopoulos, P. M., Hogenkamp, L., Gijlers, H., Khosravi, H., & Noroozi, O. (2026). Generative AI offers more, but students revise less: Comparing the effects of teacher and AI feedback on student essay revisions. *International Journal of Educational Technology in Higher Education*, *23*(1), 6. https://doi.org/10.1186/s41239-026-00579-9

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, *29*(7), 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Hacker, D., Keener, M., & Kircher, J. (2009). Writing is Applied Metacognition. In *Handbook of Metacognition in Education* (pp. 154–172). Routledge.

Kessler, M. (2021). The longitudinal development of second language writers' metacognitive genre awareness. *Journal of Second Language Writing*, *53*, 100832. https://doi.org/10.1016/j.jslw.2021.100832

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Lee, D., & Palmer, E. (2025). Prompt engineering in higher education: A systematic review to help inform curricula. *International Journal of Educational Technology in Higher Education*, *22*(1), 7. https://doi.org/10.1186/s41239-025-00503-7

Lee, I., & Mak, P. (2018). Metacognition and Metacognitive Instruction in Second Language Writing Classrooms. *TESOL Quarterly*, *52*(4), 1085–1097. https://doi.org/10.1002/tesq.436

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education: Examining research on ChatGPT's use in language learning and teaching. *Computers and Education: Artificial Intelligence*, *7*, 100266. https://doi.org/10.1016/j.caeai.2024.100266

Li, C., Lee, I., & Mao, Z. (2026). Synthesizing empirical research on learner metacognition in second language writing. *International Journal of Applied Linguistics*, ijal.70123. https://doi.org/10.1111/ijal.70123

Li, S. (2025). Generative AI and second language writing. *Digital Studies in Language and Literature*, *2*(1), 122–152. https://doi.org/10.1515/dsll-2025-0007

Mei, H., Li, L., & Liu, H. (2025). Investigating the impact of a standards-based checklist on EFL student writers' metacognitive strategy use and writing performance. *Language Awareness*, 1–20. https://doi.org/10.1080/09658416.2025.2493112

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

Nie, X., Tian, Y., Liu, M., Wu, D., & Guo, Y. (2025). The impact of generative artificial intelligence on students' higher order thinking: Evidence from a three-level meta-analysis. *Education and Information Technologies*, *30*(17), 25359–25390. https://doi.org/10.1007/s10639-025-13735-x

Qian, Y. (2025). Prompt Engineering in Education: A Systematic Review of Approaches and Educational Applications. *Journal of Educational Computing Research*, *63*(7–8), 1782–1818. https://doi.org/10.1177/07356331251365189

Qin, C., Zhang, R., & Xiao, Y. (2022). A questionnaire-based validation of metacognitive strategies in writing and their predictive effects on the writing performance of English as foreign language student writers. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.1071907

Sato, M. (2023). Metacognitive approaches to collaborative writing: Theoretical and pedagogical proposals. In M. Li & M. Zhang (Eds.), *L2 collaborative writing in diverse learning contexts* (pp. 32–52). John Benjamins Publishing Company. https://www.degruyterbrill.com/document/doi/10.1075/lllt.59.02sat/html

Sippel, L. (2024). Maximizing the benefits of peer interaction: Form-focused instruction and peer feedback training. *Language Teaching Research*, *28*(2), 413–439. https://doi.org/10.1177/13621688211004638

Teng, M. F. (2022). Effects of individual and group metacognitive prompts on tertiary-level students' metacognitive awareness and writing outcomes. *The Asia-Pacific Education Researcher*, *31*(5), 601–612. https://doi.org/10.1007/s40299-021-00611-8

Teng, M. F. (2025). Metacognition in language teaching. *Elements in Language Teaching*. https://doi.org/10.1017/9781009581295

Teng, M. F., & Ma, M. (2024). Assessing metacognition-based student feedback literacy for academic writing. *Assessing Writing*, *59*, 100811. https://doi.org/10.1016/j.asw.2024.100811

Teng, (Mark) Feng. (2020). Tertiary-level students' english writing performance and metacognitive awareness: A group metacognitive support perspective. *Scandinavian Journal of Educational Research*, *64*(4), 551–568. https://doi.org/10.1080/00313831.2019.1595712

van de Pol, J., Mercer, N., & Volman, M. (2019). Scaffolding student understanding in small-group work: Students' uptake of teacher support in subsequent small-group interaction. *Journal of The Learning Sciences*, *28*(2), 206–239. https://doi.org/10.1080/10508406.2018.1522258

Vieriu, A. M., & Petrea, G. (2025). The Impact of Artificial Intelligence (AI) on Students' Academic Development. *Education Sciences*, *15*(3), 343. https://doi.org/10.3390/educsci15030343

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3

Wang, K., Zhang, L. J., & Cooper, M. (2025). Metacognitive instruction for improving the effectiveness of collaborative writing for EFL learners' writing development. *The Asia-Pacific Education Researcher*, *34*(2), 661–673. https://doi.org/10.1007/s40299-024-00886-7

Woo, D. J., Wang, D., Yung, T., & Guo, K. (2025). Effects of a prompt engineering intervention on undergraduate students' self-efficacy, AI knowledge and prompt engineering ability: A mixed methods study. *British Educational Research Journal*, 1–28. https://doi.org/10.1002/berj.70087

Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., Zheng, C., Liu, D., Zhou, F., Huang, F., Hu, F., Ge, H., Wei, H., Lin, H., Tang, J., … Qiu, Z. (2025). *Qwen3 Technical Report* (arXiv:2505.09388). arXiv. https://doi.org/10.48550/arXiv.2505.09388

Yang, C., Shao, J., & Guo, Y. (2025). Can teacher feedback be substituted by gen-AI? A comparative study of writing scores, feedback characteristics, revision and motivation. *Innovation in Language Learning and Teaching*, 1–24. https://doi.org/10.1080/17501229.2025.2609877

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

Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*, 1–19. https://doi.org/10.1080/17501229.2025.2503890

Ziqi, C., Xinhua, Z., Qi, L., & Wei, W. (2026). L2 students' barriers in engaging with form and content-focused AI-generated feedback in revising their compositions. *Computer Assisted Language Learning*, *39*(3), 715–735. https://doi.org/10.1080/09588221.2024.2422478

Zou, S., Guo, K., Wang, J., & Liu, Y. (2025). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. *Computer Assisted Language Learning*, 1–30. https://doi.org/10.1080/09588221.2024.2447279

Supplementary Materials

Appendix S1: Overall plan and excerpted scripts of teacher modelling in the MI intervention.

Table 1. Overall arrangement of the teacher modelling session.

Components	Procedures
Introduction *(3 Minutes)*	1: Teacher highlights the pitfalls of passive AI use. 2: Establishes the goal of the session: towards activity in using GenAI.
Component 1: Goal setting *(12 Minutes)*	1: Explaining the "why" (2 mins): Teacher explains why planning is required to avoid meaningless and overly complex interactions. 2: Modeling the prompt (6 mins): Project and read aloud an exemplar prompt that sets explicit requirement and response boundaries. 3: Strategy deconstruction (3 mins): Break down how task-based, personal requests were raised and how explicit strategic instruction can keep the AI in check. 4: Student takeaway (1 mins)
Component 2: Process monitoring *(15 Minutes)*	1: Explaining the "why" (3 mins): Teacher explains how to avoid and monitor for "drifts" in context and information. 2: Modeling the prompt (7 mins): Demonstrate a tracking prompt that catches a misalignment and channeling AI back to track. 3: Strategy deconstruction (4 mins): Analyze how a conversation between AI and human helped achieve the above goal. 4: Student takeaway (1 mins)
Component 3: Evaluating quality *(15 Minutes)*	1: Explaining the "why" (3 mins): Discuss AI why and how AI might produce feedback with unsatisfactory quality. 2: Modeling the prompt (8 mins): Project a prompt where the writer skeptically challenges the major claim within AI feedback and forced AI to pay attention to the linguistic performance. 3: Strategy deconstruction (3 mins): Breaking down the prompt to show how to evaluate quality of AI feedback. 4: Student takeaway (1 min)
Component 4: Beyond one turn *(15 Minutes)*	1: Explaining the "why" (2 mins): Contrast the surface patches of single-turn prompting with the deep conceptual learning of multi-turn dialogue. 2: Modeling the prompt (9 mins): Show an interactive prompt that enables GenAI to evaluate the writing product following a user-led procedure featuring multiple rounds of iterative evaluation. 3: Strategy deconstruction (2 mins): Showing how the prompts are chained so that a continuously developing interaction pattern is maintained. 4: Student takeaway (2 mins):

Table 2. Excerpted teacher modelling script (component 1)

1. Explaining the “why”

[Teacher explanation]: Before we start using the AI, we must focus on the ***planning*** stage of metacognition. When students ask an AI for 'feedback' without a specific goal, the AI usually provides corrections on grammar or spelling. However, effective writing involves more than just being correct.

We have to do relatively more than expected with AI feedback process. So, we start with getting feedback. The better we get initially, the less effort is demanded for following steps. So, we must first think about the **task we are facing** [add some explanation if needed] and yourself as a ***person*** [add some explanation if needed]. If we do not set a goal first, we will receive a large amount of feedback that may not address our actual needs. Planning allows us to limit the AI's response to the specific areas we want to improve.

2. Modeling the prompt

[Teacher introduction]: I will show you a prompt I have prepared. Notice how this prompt includes specific information about the person, the task, and the strategy I want the AI to use.

[Showing the task and sample writing]

(omitted)…

[Projected on screen]:

*I am writing an argumentative essay about the environmental impact of fast fashion. This is for a university-level English course, and the tone must be academic. Here is the rating rubric of the writing task [omitted] (**task**). I know that I often write sentences that are too long and wordy, which makes my ideas hard to follow, so I am expecting you to act as an EFL teacher good at this point (**person**). I am not looking for grammar corrections right now. Instead, please look at my second paragraph and highlight three sentences that are too complex. Explain why they are difficult to read so that I can rewrite them myself. Here is a template for you to follow, which is in alignment with the rubric [omitted] (**strategy**).*

[Teacher explanation]: This prompt is an example of the **planning** strategy. I am not just asking the AI to fix my paper. I am telling the AI exactly what my problem is and how I want it to help me solve it.

3. Strategy deconstruction

[Teacher explanation]: Let’s break down how this prompt connects to our metacognitive framework. We are using what we know to guide the feedback process.

Prompt element	**Metacognition knowledge**	**Prompt engineering techniques**	**Practical function**
"I write sentences that are too long..."	Person	Persona	You identify your personal weakness, and your expectations of characteristic of AI as a human-like evaluator.
"University-level... academic tone..."	Task	Reference texts	You tell the AI the required standard for the assignment and add grading rubric so that its output is reference-based.
"Explain why... so I can rewrite them..."	Strategy Knowledge	Template	Give very specific requirements regarding the function and the presentation of AI output.

Note: All the prompt engineering techniques were based on Woo et al. (2025).

4. Student takeaway: advice for use

[Teacher advice]: Think before prompting. Answer questions about tasks, person, and strategy then complete your prompt. By doing so, the feedback you receive will be much more relevant to your needs.

Appendix S2: Sample of think-aloud utterances on feedback uptake.

Category	GenAI feedback*	Think-aloud utterance	Feedback modification (if any)
Ignore	Consider exploring the inherent power dynamics in the teacher-student relationship. Discussing how students navigate boundaries can deepen your analysis of authority structures.	Honestly, [the feedback] sounds way too formal for my essay. I'm just trying to write about how to talk to a teacher when you're stressed.	[Feedback rejected]
Elaboration / modification	To make your point about "showing respect" more concrete, provide an example of active listening. For instance, you could describe a student nodding and taking notes during a lecture.	I don't feel like it, it's too "old school" … Instead, I modify this advice to focus on a student asking to clarify questions during office hours. That still proves the concept of active listening but feels at least modern.	Students keep the general suggestions (e.g., "provide a concrete example of respect") but replaces the micro-scaffolding (nodding/notes' with "office hour dialogue").
Apply	The transition between paragraph 2 (respect) and paragraph 3 (grading conflicts) is not natural. Try a transitional phrase like: "However, this mutual respect is often put to the test when grading conflicts arise."	… it points out that my transition between the two parts is truly not good enough. AI feedback makes sense…	Students accept the critique with very minimal language modification (i.e., reducing "often put to the test" to "tested").

Note: *: Excerpted from AI-student chat history.

Appendix S3: Codebook for qualitative journal analysis.

Themes and codes	"Sources"		Frequencies		Knowledge/regulation	Samples
	CG	EG	CG	EG		
1. Making plans						
1.1 Setting persona / specify AI roles	5	13	12	21	Strategy	"Act as a tutor..." "I want you to act as if an experienced English teacher..."
1.2 Incorporating personal preferences/needs	12	13	15	14	Person	"I need simple explanations because I'm a beginner."
1.3 Incorporating task-specific info	41	50	63	77	Task	"The assignment needs 3 arguments with evidence." "This is a college-level essay task, so focus on having more insights." "The process requires step-by-step reasoning."
1.4 Pre-interaction strategy choice	48	57	86	92	Strategy	"I'll use the AI for initial brainstorming then do my own revision." "I'm starting with questions to activate my prior knowledge."
2. Controlling quality of AI feedback						
2.1 Accuracy verification	11	43	15	68	Evaluating + monitoring	"The AI's logic seems to contradict the exemplars we have." "It's counter-intuitive… so I began checking it's right or not."
2.2 Relevance & depth judgment	8	36	11	54	Evaluating + strategy	"The AI drifted away from the requirements I provided."
2.3 Identifying AI limitations	7	31	8	47	Evaluating + monitoring	"The AI fabricated a source that doesn't exist."
2.4 Feedback evaluation & iteration	12	54	14	72	Evaluating + monitoring	"I asked for more concrete examples because the first ones were too vague."
3. Navigating conversation with AI						
3.1 Prompt adaptation	13	56	17	81	Task + strategy + monitoring	"I summarized our previous exchange to get the conversation back on track."
3.2 Conversation monitoring	5	43	9	65	Monitoring	"I urge you to offer suggestions alongside the original content, preferably in the form of a table."
3.3 Adaptive strategy regulation	10	44	10	58	Monitoring + strategy	"The broad explanation didn't help, so I switched to asking for a real-world example."
3.4 Persistence & decision-making	9	41	13	49	Monitoring + evaluating	"I kept iterating until I finally understood the suggestion." "The AI was becoming unproductive and repetitive, so I decided to disengage and use the library instead."