Accepted for publication in Assessment and Evaluation in Higher Education on 1 August 2025

Designing ChatGPT-Mediated Feedback Activities in EFL Writing: A Design-Based Study of the Dialogic Feedback Triangle

Yunan Zhang

u3010204@connect.hku.hk

[1]Faculty of Education, University of Hong Kong

[2]Faculty of Education, University of Cambridge

Yongcan Liu (Corresponding author)

yl258@cam.ac.uk

Faculty of Education, University of Cambridge

Abstract

With the rise of GenAI-assisted English as a Foreign Language (EFL) writing, chatbots such as ChatGPT have shown potential both as formative feedback providers and interactive partners, offering new possibilities for dialogic feedback. However, little research has explored how educational practitioners can design and implement ChatGPT-mediated dialogic feedback activities. Addressing this gap, this study adopted a design-based research (DBR) approach to iteratively develop an optimal design for ChatGPT-mediated dialogic feedback activities, drawing on a dialogic feedback triangle as a conceptual framework. Across three iterative cycles with four EFL learners, data were collected through semi-structured observations and interviews focused on participants' writing experiences and feedback on the intervention. The three research rounds led to the identification of seven key design elements, organised into a three-dimensional feedback triangle: cognitive (disciplinary knowledge, cue-consciousness, self-evaluation), socio-affective (adaptive roles of ChatGPT at different writing stages, learners' critical attitudes towards ChatGPT), and structural (mobilisation of learners' tool repertoires, feedback structures based on specific formats and standards). These design elements can offer insights into feedback literacy and inform seven practical deisign principles to guide educators and technology designers in designing and implementing effective GenAI-mediated dialogic feedback experiences in EFL writing contexts.

Introduction

Recent years have seen increasing research interest in Generative AI (GenAI)-assisted English as a Foreign Language (EFL) writing (Barrot, 2023). GenAI chatbots have been employed to provide feedback for writers, offering ratings with explanations (Mizumoto & Eguchi, 2023) and evaluating accuracy (Poole & Coss, 2024). However, current GenAI-assisted EFL writing applications primarily offer detailed evaluations based on predetermined criteria. This approach mirrors traditional monologic feedback, where feedback is treated as a one-way transmission of information, arriving only at the final assessment of students' work (Carless et al., 2011).

Dialogic feedback presents an alternative approach where students interact and co-construct their understanding rather than passively receiving feedback (Ajjawi & Boud, 2017). Dialogic feedback encourages students to critically evaluate and make decisions regarding the received feedback, fostering engagement and learning gains (Carless et al., 2011). Despite its benefits, implementing dialogic feedback in teacher- or peer-mediated contexts can meet logistical challenges due to time limitations and large class sizes (Nicol & Macfarlane-Dick, 2006; Carless, 2016). GenAI chatbots offer a promising opportunity to support the design and delivery of dialogic feedback by functioning as both evaluators providing timely formative feedback and interactive partners engaging with students (Su et al., 2023). However, limited research has explored an optimal way for educational practitioners to design ChatGPT-mediated dialogic feedback activities (Zhang & Liu, accepted). As such, pedagogical guidance to support teachers or educational technology designers in integrating ChatGPT into dialogic feedback practices remains underdeveloped.

Addressing this gap, this research has iteratively developed an optimal design for ChatGPT-mediated dialogic feedback activities through a design-based research (DBR) approach. The final design elements have been structured around a dialogic feedback triangle encompassing the cognitive, socio-affective, and structural dimensions, providing insights into feedback literacy. Practical implications were derived from the feedback triangle to guide future educators and educational technology designers in designing and implementing GenAI-mediated dialogic feedback activities.

Literature Review

Dialogic Feedback and Feedback Triangle

Although feedback is vital to learning, it remains one of the most unsatisfactory aspects of university education, partly due to its one-way, teacher-driven design and delivery (Nicol & Macfarlane-Dick, 2006; Carless, 2013; Wood, 2021). Such monologic feedback practices optimistically assume that students can independently understand and apply feedback to future tasks (Carless, 2013). Positioned as passive learners, students often struggle to decode the feedback, missing its potential benefits

(Sutton & Gill, 2010). Therefore, it is the strategic design and implementation of feedback activities that largely influence students' engagement and the effectiveness of feedback, rather than feedback itself (Ajjawi & Boud, 2017; Wood, 2021).

Given the limitations of monologic feedback, researchers have called for a shift towards dialogic feedback to engage learners in a process to actively co-construct understanding through discussion and reflection (Yang & Carless, 2013; Erkan et al., 2021; Wood, 2021). This approach aligns with socio-constructivist learning theories where knowledge is viewed as socially negotiated rather than passively received (Carless et al., 2011). Empirical studies demonstrate the advantages of dialogic feedback such as fostering deeper engagement with learning materials (Bloxham & Campbell, 2010), helping students interpret comments and make evaluative judgments (Zhu & Carless, 2018), and promoting self-regulated learning (Ajjawi & Boud, 2017.

Despite these benefits, there is a gap between theory and practice in how feedback dialogues are designed and employed, calling for optimistic designs and guidance for participants (Wood, 2021). To operationalise dialogic feedback, Yang and Carless (2013) proposed a dialogic feedback triangle (Figure 1) that addresses three main barriers to feedback practices and structures feedback into three interconnected dimensions: the cognitive, focusing on feedback content; the social-affective, emphasising interpersonal negotiation; and the structural, addressing the organisation of feedback delivery (Yang & Carless, 2013). Each dimension contains factors that can either support or inhibit effective feedback, with the interaction among these elements playing a crucial role (Yang & Carless, 2013). Based on Figure 1, six features to construct dialogic feedback environments have been derived, as shown in Figure 2.

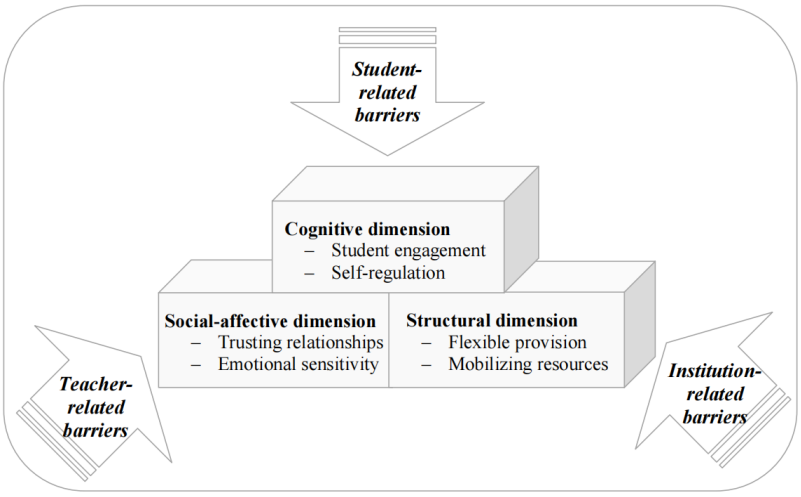

Figure 1. The Architecture of Dialogic Feedback (Yang & Carless, 2013, p.293)

Six features to construct dialogic feedback environments:
(1) stimulating student engagement with disciplinary problems through dialogic feedback;
(2) developing student self-regulation by helping them understand feedback purposes and learner roles;
(3) nurturing collaborative and trusting relationships;
(4) attending to students' emotional and psychological needs;
(5) offering flexible feedback in timing, form, and sequencing;
(6) mobilising both disciplinary and non-disciplinary resources, especially new technologies.

Figure 2. Six features to construct dialogic feedback environments (Yang & Carless, 2013)

Overall, the dialogic feedback triangle provides a comprehensive perspective to analyse and optimise the elements of a feedback activity and their interactions. However, the triangle mainly attends to teacher-led dialogic feedback, while research is scarce on applying this model to other contexts, especially technology-enhanced settings.

Technology-Enhanced Dialogic Feedback

Carless (2016) identifies five key ways to operate the dialogic feedback processes: teacher guidance, peer review, technology-mediated dialogues, students' internal feedback, and written teacher feedback informed by students' expectations in assignments. Among these, technology-enhanced dialogic feedback has attracted increasing attention, delivered through audio (Gould & Day, 2013), video (Hidson, 2024), screencasts (Wood, 2023), and written feedback via digital platforms including collaborative document platforms (Andrade & Roshay, 2024) and e-portfolio systems (Hui et al., 2024). Such feedback can overcome the temporal, spatial, and resource constraints of face-to-face interactions, enriching learner engagement and feedback uptake (Wood, 2021). These feedback also shows promising benefits in cognitive and affective aspects, through preserving the digital records of dialogues to support reflections (Wegerif, 2013) and fostering socio-affective engagement and teacher-learner relationship (Wood, 2023).

However, current technology-enhanced dialogic feedback faces two key challenges. First, sustaining student engagement remains difficult. Many practices are designed to support asynchronous exchanges, where delayed responses between assessors and assessees could hinder the flow and interactivity of dialogues (Van Hoe et al., 2024). Formats like audio, video, or screencast feedback often lack convenient channels for student responses (Wood, 2023). Thus, without a "feed-forward" mechanism to promote ongoing exchanges, the dialogues risk stalling at one-way delivery (Dowden et al. 2013; Andrade & Roshay, 2024). Even platforms with back-and-forth commenting may not sustain interaction, as students often stop engaging after the first comment (Hui et al., 2024), limiting active engagement and feedback uptake (Van Hoe et al., 2024). Second, a broader challenge lies in the logistical constraints of human-provided feedback, which is no unique to

technology-mediated feedback but also applies to face-to-face peer and teacher dialogues. The five feedback approaches identified by Carless (2016) depend on teachers or peers, and even digital platforms still rely on human input. However, teacher-led dialogic feedback is often constrained by time, class size, and institutional resources (Nicol & Macfarlane-Dick, 2006; Carless, 2016). Similarly, peer feedback practices can meet challenges such as variable quality, coordination difficulties, and students' reluctance to engage critically (Carless, 2016).

Altogether, these challenges point to the need for alternative mechanisms that can both promote interactive dialogues and student engagement, as well as address the logistical burdens in human-mediated feedback. We then proposed GenAI-mediated feedback as a potential approach, which will be illustrated in the next section.

GenAI-Mediated Feedback

GenAI chatbots offer a potential solution to the aforementioned challenges through providing immediate and interactive feedback (Huang et al., 2022). GenAI chatbots can generate context-relevant and human-like responses, acting both as an evaluator and conversational partner to provide immediate formative feedback and address follow-up questions (Barrot, 2023; Su et al., 2023). In this study, we conceptualise GenAI chatbots not simply as tools for general writing assistance, but more specifically as potential mediators of dialogic feedback. Building on Yang and Carless's (2013) feedback triangle (Figure 1), GenAI chatbots could facilitate dialogic feedback cognitively, social-affectively, and structurally. Cognitively, GenAI chatbots allow students to engage in real-time adaptive feedback cycles (Su et al., 2023), inviting more investment of cognitive efforts and potentially promoting engagement compared to the aforementioned asynchronous feedback. Social-affectively, chatbots may help mitigate social barriers in peer review, as students may feel more comfortable while engaging in trial-and-error learning without fear of judgment (Huang et al., 2022). Researchers also found reduced anxiety and enhanced willingness to communicate when learners use GenAI chatbots (Teng, 2024), which could help solve the challenge on student engagement in dialogues. Structurally, chatbots provide scalable and flexible feedback, overcoming constraints such as large class sizes and limited teacher availability (Carless, 2016), thus sustaining interactive feedback across large cohorts. Accordingly, these affordances suggest that GenAI chatbots have the potential to address two aforementioned challenges in technology-enhanced dialogic feedback on student engagement and logistical constraints of human-mediated feedback provision.

Meanwhile, GenAI chatbots could introduce new constraints that challenge the established six assumptions of effective feedback practice in Yang and Carless (2013). Cognitively, ChatGPT often produce excessive feedback covering every dimension in the assessment criteria (Guo & Wang, 2023; Lee et al., 2025), potentially posing cognitive challenges for information processing. Then, while the social-affective dimension emphasises mutual trust, GenAI cannot replicate the interpersonal aspect of human feedback (Hu et al., 2024). Rather, effective use of GenAI may not involve mutual trust, given the concerns regarding inaccurate information caused by artificial

hallucination and biased content (Alkaissi & McFarlane, 2023; Guo & Wang, 2023; Foung et al., 2024). From the structural dimension, relying on generalised rather than domain-specific training data, GenAI chatbots may fall short on the sixth feature on mobilising both disciplinary and non-disciplinary resources (Hu et al., 2024). GenAI chatbots thus cannot apply disciplinary-specific resources as effectively as a teacher does as aforementioned (Foung et al., 2024). Altogether, the previous established features of effective feedback as envisioned by Yang and Carless (2013) may not fully apply in GenAI-mediated settings given the limitations of GenAI chatbots. This further highlights the need for reconstructing the design features of feedback to align with the GenAI context.

In sum, one main driver motivates the use of GenAI chatbots in dialogic feedback activity: these GenAI tools could help address challenges faced by previous human-mediated and technology-enhanced feedback and offer new possibilities for enhancing feedback across cognitive, social-affective, and structural dimensions. The unique affordances and limitations of GenAI chatbots further call for a need to rethink the dialogic feedback design beyond the previously established features in Yang and Carless (2013). To the best of our knowledge, no research has explored the potential of GenAI chatbots through the lens of dialogic feedback. There has been some research on writers interacting with chatbots, while researchers may optimistically adopt a naturalistic approach and rely on the potential that optimal feedback practices can emerge without intervention (Zhang & Liu, Accepted). Without a design-based interventionist approach, these practices may lack the support necessary to guide learners in strategically using AI (Yan, 2023). This could be particularly relevant for ChatGPT-mediated writing, given the limitations of GenAI chatbots and concerns over plagiarism (Zhang & Liu, Accepted). These challenges highlight the need to re-design dialogic feedback activities through a DBR approach that can enable the iterative development of a pedagogically grounded and practically viable intervention (Akker, 2006). Developing design principles for such intervention can thus inform teachers' professional practices and educational technology developers to design optimal ChatGPT-mediated feedback activities.

The Present Study

This research drew on the dialogic feedback triangle to conceptualise and develop an optimal design of ChatGPT-mediated dialogic feedback activities through a DBR approach. The research question is: how a ChatGPT-mediated feedback activity can be constructed to foster effective dialogues? Cognitive, social-affective, and structural elements of optimal ChatGPT-mediated dialogic feedback practices have been identified and modeled into a three-dimensional and seven-element feedback triangle. We finally drew out seven practical implications from the triangle to guide future practitioners' design and implementation of ChatGPT-mediated dialogic feedback activities.

Methods

DBR was adopted with a qualitative multiple-case study design. DBR is a paradigm "for the study of learning in context through the systematic design and study of instructional strategies and tools" (The Design-Based Research Collective, 2003, p.1). DBR is suitable for this research given DBR's interventionist and iterative nature, duality in theory and practice, and researcher-practitioner collaboration (The Design-Based Research Collective, 2003). The iterative cycles of design, implementation, and revision in DBR allow for continuous refinement and improvement of interventions (Bakker, 2018), specifically, the ChatGPT-mediated dialogic feedback activity design in this research. Regarding DBR's duality in theory and practice, empirically, DBR can develop design elements to integrate AI into actual contexts and theoretically develop a dialogic feedback triangle to understand the feedback activities in-depth. In DBR, practitioners are not passive subjects of intervention but co-designers bringing in their practical knowledge and contextual understanding to develop interventions (Bakker, 2018). This can fully incorporate student writers' voices into the development of the intervention.

Research Context and Sampling

This study was conducted in a Chinese EFL context, using IELTS Writing Task 2 as a representative argumentative writing task. Three tasks from the official test book *IELTS 17* were selected. Four Chinese EFL students were recruited based on the following criteria: (1) access to ChatGPT 3.5 and willingness to use it in EFL writing; (2) sufficient English proficiency to ensure interactions with ChatGPT 3.5, which performs best in English (OpenAI, 2024); (3) prior experience in EFL argumentative writing to support the participation; and (4) varying levels of familiarity with ChatGPT, allowing for a diverse dataset. We summarised participant profiles in Figure 3 (see also Zhang & Liu, Accepted):

A Summary of Participant Profiles

(1) Wang had taken the IELTS writing test twice and achieved a score of 7. Prior to the study, she was unfamiliar with ChatGPT and rarely used it.

(2) Yan had also taken the IELTS writing test twice and achieved a score of 7. Seldom prior experience with ChatGPT was reported.

(3) Zhang had taken the IELTS writing test once and received a 7.5 in writing. With a background in linguistics and a Master's-level course in computational linguistics, Zhang frequently used ChatGPT for academic reading and writing purposes.

(4) Li had scored 25 in TOEFL writing and was preparing for the IELTS writing test. Holding a PhD in Natural Language Processing (NLP), he regularly used ChatGPT in his everyday writing.

Figure 3. Participant Profiles

Research Procedure

As visualised in Figure 4, the research was conducted through the three main phases of a typical DBR design according to Bakker (2018): preparation and design,

intervention and enactment, and retrospective analysis.

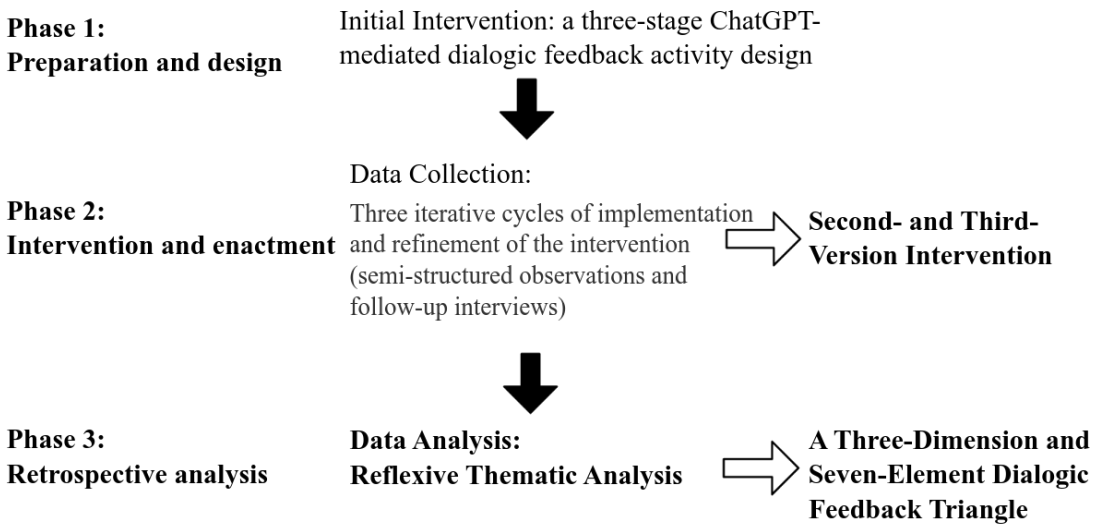

Figure 4. Research Procedure

First, DBR begins by identifying a practical problem (Bakker, 2018), which, in this research, is the RQ informed by the literature review. As shown in Figure 4, in Phase 1, we developed an initial version of the intervention, a three-stage ChatGPT-mediated dialogic feedback activity design based on Su et al. (2023). Based on their early experimentation with ChatGPT, Su et al. (2023) provided concrete suggestions and example prompts for using the tool across key stages of EFL argumentative writing, including outlining, drafting, and proofreading. We adapted these suggestions to suit the IELTS writing context, developing an initial intervention that was hypothetical and subject to further revision. For example, Figure 4 is a screenshot of the initial intervention on instructing writers to use ChatGPT for outline generation. In Figure 5, writers are advised to write an initial outline independently, use the prompt to ask ChatGPT to provide feedback, and engage in following dialogues with ChatGPT.

Stage 1: Outline Generation

- At this stage, you should write an outline for the task first on your own and then use ChatGPT to seek feedback and additional ideas.
- While writing your outline, please follow a CER (claim-evidence-reasoning) framework (McNeil & Krajcik, 2009).

After writing your outline, you can send this prompt to ChatGPT:

"Act as an argumentation outliner who is currently working on providing feedback to the outline of EFL students' IELTS task 2, i.e., providing critical, specific, and constructive feedback to the outline of the writing.

Please follow the checklists to evaluate whether the outline has incorporated the essential elements of argumentative writing and whether the outline can address the topic well. Provide critical and specific feedback to the outline with some explanations.

Overall Outline Evaluation Criteria (below is an example for whoever target at score 7. Please refer to the task response and coherence and cohesion sections of your target score band descriptors) - addresses all parts of the task - presents a clear position throughout the outline - logically organises information and ideas; there is clear progression throughout

Claim and sub-claims: - the outline has a main claim that is relevant and well-suited to the topic; - the sub-claims are closely connected and progressed logically to support the main claim

Reasoning:- the reasoning provides a justification of why the evidence support the claim;

Evidence:- the evidence can support the sub-claims; - the evidence is from credible and trustworthy sources;

I will give you an outline later. Do not start until I provide you with it. Please be specific. Do you understand?"

- You are encouraged to ask follow-up questions to ChatGPT to revise your outline

Figure 5. Screenshot of the Initial Intervention on Instructing Writers to Use ChatGPT for Feedback on their Outlines

Phase 2 involved three iterative cycles of implementation and refinement with the participants, with each cycle generating new data to improve the intervention and leading to second- and third- version interventions (Figure 4). Within each cycle, semi-structured observations were conducted with each participant on their writing experience and feedback on the intervention design. In the interviews, we asked

questions such as “How do you feel about the writing guideline in general?” and “Do you have suggestions for improving this part” which served as the starting point to gather participants’ experiences and perceptions. These led to revisions of our intervention (See Appendix A for the evolution of the intervention). The research was conducted via online meeting software and in their chosen languages (English for Yan and Mandarin for Wang, Zhang, and Li, each lasting 30 to 45 minutes).

In Phase 3, writers’ chat logs and the four versions of the intervention were collected as data triangulation and analysed through reflexive thematic analysis (RTA). RTA is “a theoretically flexible method” for “developing, analysing, and interpreting patterns across a qualitative data set” while incorporating theoretical perspectives (Braun & Clarke, 2021, p.4). Echoing some insights on reflexivity in reflexive discourse analysis (Alejandro, 2020), researcher reflexivity in this study was framed as how our personal experiences, theoretical perspectives, and discursive choices shape the knowledge construction. The interview data were analysed via NVivo using RTA based on the dialogic feedback triangle, following a six-stage coding process based on Braun and Clarke (2021): familiarisation, transcription, generating initial descriptive codes, grouping, refining and defining the themes. The coding process was both inductive, using themes from the dialogic feedback triangle, and deductive, with sub-themes developing throughout the coding. Finally, the feedback design was modeled as a feedback triangle.

Findings

We conducted three research rounds to iteratively revise the intervention (i.e. the ChatGPT-mediated dialogic feedback activity design) underpinned by the feedback triangle. In Iteration 1, major revisions were made regarding our intervention. Then, participants found the second-version intervention significantly more useful but still pointed out some problems, so minor refinements were made in Iteration 2. In Iteration 3, participants widely showed positive attitudes and had few suggestions, which justified the need to conclude the research. The specific revisions across the three cycles will be explained later in this section. Appendix A presents these revisions and screenshots of the revised intervention with corresponding design elements and recommendations for relevant activity design. Figure 5 operationalises our findings into a three-dimension and seven-element triangle of the optimal dialogic feedback activities. These seven elements have emerged from our three iterations of research and been organised around the three interconnected dimensions. Based on Yang and Carless’s (2013) framework for effective feedback, the cognitive dimension refers to students’ cognitive attributes that should be developed through feedback. In this study, these attributes include disciplinary knowledge, cue-consciousness, and self-evaluation. Concerning the social-affective dimension, namely, the relationship between feedback providers and receivers, our design highlights the adaptive roles of ChatGPT at different writing stages and students’ critical attitudes towards ChatGPT. For the structural dimension on the organisation of feedback and deployment of resources, two elements have been identified: the mobilisation of learners’ tool

repertoires and feedback structures based on specific formats and standards. The feedback triangle in Figure 6 shows the dynamic interplay among these dimensions in a holistic feedback space. Though each of these elements has been posited within the most relevant dimension, they should be best understood within the interplay of cognitive, social-affective, and structural dimensions.

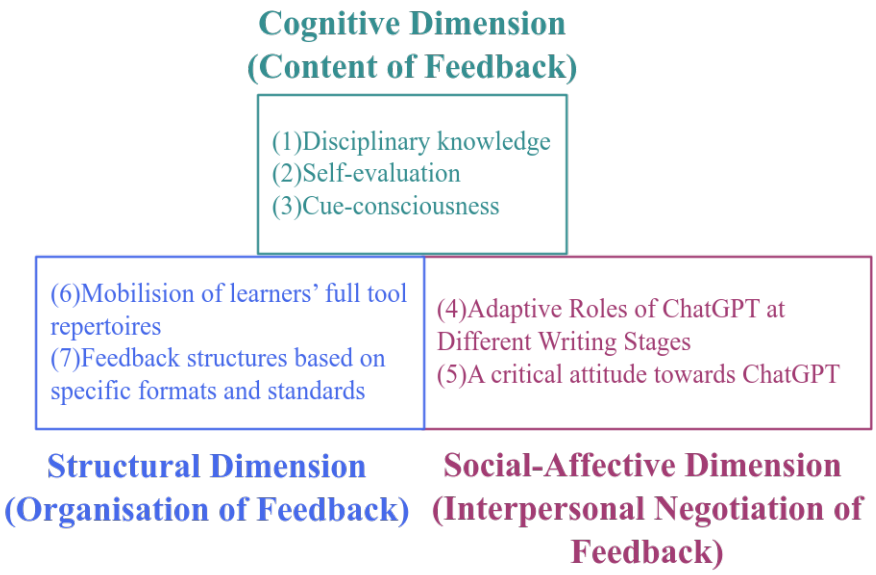

Figure 6. A Three-Dimension and Seven-Element Triangle of the Optimal ChatGPT-Mediated Dialogic Feedback Activity

Organised by the three dimensions, this section will present each of these design elements, how they emerge from the research and inform the design principles of the feedback activity, and how the revised design could afford dialogic feedback. In each round, we made revisions to our intervention according to participants' feedback and perceptions. The design principles were then derived from the process, underpinned by the design elements.

Cognitive Dimension (Content of Feedback)

Disciplinary knowledge

The first cognitive attribute that the feedback activity aimed to foster was students' understanding of concepts within a specific discipline and skills for investigating disciplinary problems. The corresponding design principle was identified as "providing discipline-specific guidance to foster disciplinary knowledge." During Iteration 2, Wang expressed a limitation of ChatGPT, noting that its output lacked the specificity and domain expertise of IELTS teaching videos she watched before. She remarked as below:

"ChatGPT may be more suitable to whom has a certain foundation, otherwise the effect should not be as good as the small (teaching) videos I have seen to tell me how

to write according to each question type."

According to her, without prior knowledge of IELTS writing, learners may not find the ChatGPT-mediated writing helpful in learning. Her observation highlights a critical cognitive dimension: the need for domain-specific guidance in understanding and applying IELTS writing structures. The lack of such tailored guidance may constrain the effectiveness of the feedback activity, showing the importance of disciplinary knowledge. To address this constraint, we consulted relevant learning materials on IELTS writing and added some guidance on writing IELTS essays to the intervention, aiming to foster writers' disciplinary knowledge.

Self-evaluation

Self-evaluation, as an attribute of self-regulative learners, was spotlighted to encourage writers to self-monitor their work and independently develop ideas. The corresponding design principle was "promoting students' self-evaluation of their work before turning to ChatGPT." In the initial design, participants were encouraged to engage with ChatGPT throughout the writing process. However, this approach sometimes resulted in over-reliance on ChatGPT particularly when participants encountered challenging topics. For example, Wang noted that she increased her reliance on ChatGPT during Iteration 3, as the writing task was particularly difficult for her. She described this change in her relationship with ChatGPT as "power dynamics," explaining how her own agency decreased while ChatGPT's role became dominant:

"The first two times... I was very proactive and had some of my own opinions. My power was very strong, so its (ChatGPT's) power was relatively weak... Then, this time my power was very weak because I had no idea... Its power was like a spring. It would be stronger... I needed to be supported completely by it."

As illustrated, the lack of self-evaluation of the writing task and one's writing made Wang more vulnerable to over-reliance when she was not familiar with the task, potentially constraining her independent thinking. Thus, encouraging self-evaluation before introducing ChatGPT's support could help address this challenge. The revised design incorporated explicit reminders at two stages: During outline generation, participants were encouraged to first brainstorm and document their ideas before consulting ChatGPT. During content drafting, an additional note emphasised the importance of maintaining agency and avoiding direct copying of ChatGPT's output. When writers take the first step independently, they are expected to maintain some cognitive ownership of their ideas, even when later collaborating with ChatGPT.

Cue-consciousness

Cue-consciousness is a critical attribute that involves identifying signals in tutors' discourse about the discipline and assessment requirements, and how to achieve optimal results (Boekaerts, 2010). Students' cue-consciousness in this study is

fostered through providing graduated mediation from implicit to explicit, with a scoring element. Specifically, the proofreading stage was structured into two stages, one for general and implicit feedback and the other for focused and explicit instructions, and ChatGPT provided scoring to provoke writers' reflections. This design principle, with a focus on assessment criteria, can heighten students' sensitivity to domain-specific requirements in a graduated manner, promoting cue-consciousness (Nicol & Macfarlane-Dick, 2006).

In the initial design, when participants used ChatGPT for proofreading, ChatGPT often generated overly general feedback, which participants deemed largely unhelpful. They were cue-deaf at this stage, failing to make use of the feedback due to its vagueness. While interviewees did not provide explicit suggestions to address this issue, they were observed using a shared strategy to counteract ChatGPT's vagueness: asking it to elaborate on specific writing dimensions that need improvement. After Iteration 1, we made a methodological decision to derive possible solutions from literature: incorporate graduated mediation - from implicit to explicit feedback - and add a scoring element. Specifically, indirect and implicit feedback that aims to raise learners' attention without inclusion of the correct form can benefit learners' long-term development, while direct and explicit feedback provides clearer and more useful information (Ferris et al., 2013). This indicates that a two-phase process - starting with implicit evaluation and progressing to explicit prompts can be valuable. Moreover, ChatGPT has been used as an automated essay scoring tool and shown some degree of validity and reliability (Mizumoto & Eguchi, 2023). A scoring element respectively on four IELTS writing dimensions was then added to promote writers' reflections, with a reminder of ChatGPT's limited reliability.

In Iteration 2, this change was found to enhance both cognitive and affective engagement of participants, showing the revised design's utility. According to Wang, cognitively, the scoring element heightened her awareness of writing standards and problem areas, as said by her "If the score is exceptionally high or low, it may attract me to carefully look at why it is so." Similarly, Yan highlighted that scoring could serve as a motivational tool for novice writers, contributing to the social-affective dimension of the feedback activities. Altogether, the adjustment of the prompts for proofreading led to a significant change in feedback content, enhancing writers' cue-consciousness of IELTS writing.

Socio-Affective Dimension (Interpersonal Negotiation of Feedback)

Adaptive Roles of ChatGPT at Different Writing Stages

The design element on adapting ChatGPT's roles to different writing stages aims to tailor ChatGPT to specific writing needs, positioning ChatGPT as a brainstorming peer before writing and professional assessor after writing. In the initial intervention, ChatGPT was positioned merely as a feedback provider. However, as aforementioned, our revised design emphasised differentiated strategies for each stage, such as brainstorming with ChatGPT before writing and gaining both explicit and implicit feedback with a scoring element after writing. These changes redefined ChatGPT's role, allowing it to shift dynamically from a peer collaborator during brainstorming to

an assessor-like role in the revising phase. Yan reflected on this progression, stating that:
"For the first part, like the earlier stages, it's more like a friend or co-writing peer because I have my own ideas. It has its own ideas. We write it down, compare it together, so it's more like peer work. But then towards the end, it's more like a teacher or instructor, assessor...because it gave me feedback to improve the grade."

As illustrated by Yan, cognitively, ChatGPT's peer-like role at brainstorming stage helped her expand ideas, while the assessor-like role in later stages focused on revising languages and structures. Social-affectively, such adaptability of ChatGPT could encourage Yan to engage deeper in the interactions, fostering a sense of partnership early on and guidance later. These dialogic negotiations with ChatGPT by aligning its role with different writing stages were perceived as effective by participants. By aligning ChatGPT's interventions with distinct writing stages, this design element could foster a context-specific co-regulation with ChatGPT as writers move from generating ideas to revising their final drafts.

A Critical Attitude towards ChatGPT

This feature aims to foster a critical attitude towards ChatGPT's suggestions while motivating writers to engage in iterative dialogues with ChatGPT. During Iteration 1, Zhang viewed ChatGPT as an authoritative source for writing advice, admitting that he "kind of blindly believed" suggestions from ChatGPT without verifying their accuracy. However, at the later proofreading stage, ChatGPT flagged a grammatically correct sentence as erroneous and offered a revision that changed its intended meaning. Zhang perceived such overcorrection as "erroneous" and misleading feedback, expressing his frustration: "I found out that its advice was wrong, and I was a little annoyed." This negative affection significantly changed Zhang's perspective of ChatGPT in the second iteration. Reflecting on his earlier disappointment, he adopted a more critical view, describing ChatGPT as "a tool to be used" rather than an authority. This critical attitude promoted him to scrutinise ChatGPT's output, saying "Because the last time it lied to me, I remember that this time." This progression from blind trust to critical engagement shows the importance of maintaining a critical stance. Social-affectively, Zhang reduced his trust in ChatGPT and showed some degree of frustration and annoyance, which made him redefine his expectations of ChatGPT as a non-human tool. Structurally, his critical attitude creates some space for dialogues with ChatGPT through multiple rounds of prompts and responses. Altogether, with a critical view on ChatGPT's feedback, Zhang went beyond superficial uptake of feedback and engaged in deep learning.

To encourage a critical approach, the revised design incorporates some guidance on ChatGPT's limitations at every stage: writers are advised that ChatGPT may provide vague or inconsistent comments and are encouraged to run their writing through ChatGPT multiple times and cross-check the consistency of feedback. Meanwhile, writers are encouraged to ask follow-up questions for clarification. These changes aim to promote participants' critical engagement with ChatGPT in their

dialogues.

Structural Dimension (Organisation of Feedback)
Mobilisation of Learners' Full Tool Repertoire
This element is to encourage writers to use a range of tools alongside ChatGPT, to exert their complementary roles for ChatGPT. Although the initial guideline articulated that writers were encouraged to use all available resources, in the first iteration, Zhang expressed that ChatGPT was sufficient to assist the writing, with little need to use other tools. By contrast, Yan consistently utilised Microsoft Word's automatic typing function across the three research rounds and found it valuable:
"I think they (ChatGPT and Microsoft Word automatic typing) are quite similar, so if I have time, I would go through both and then have more diverse opinions...you can compare both before you decide what or where to improve."

By framing ChatGPT and Microsoft as parallel sources of feedback, Yan emphasised the importance of comparing their suggestions to inform her decision-making, reflecting her critical engagement with the two tools. The use of a full tool repertoire for cross-referencing feedback could potentially balance the reliance on AI-generated input with writers' independent judgment. Other participants also demonstrated a proactive approach to tool usage. For example, Wang consulted an e-dictionary for authoritative vocabulary examples, preferring it over ChatGPT for authentic language use because she "can see how authoritative it (sample sentence) is and then see how the sample sentence is used." Additionally, Li used search engines to confirm evidence for his arguments, which together highlight the complementary roles of other tools alongside ChatGPT. Thus, structurally, integrating multiple tools can foster a more dynamic feedback mechanism through using writers' whole tool repertoire. The intervention thus explicitly incorporates recommendations to use complementary tools including e-dictionaries, search engines, and Grammarly.

Feedback Structures based on Specific Formats and Standards
This feature aims to ensure ChatGPT generate specific feedback in predefined formats and criteria through specifying prompts. In the original intervention, Li encountered limitations when using ChatGPT to generate an outline. He noted that the evidence provided by ChatGPT was "just paraphrasing, very vague, and not useful." Thus, ChatGPT failed to meet his expectation of providing useful examples that could support his arguments, leading to an unproductive dialogue. He suggested, "We could tell it, for example, to give examples of real person or companies in the real world, tell it to be more detailed and more specific, which may be useful." Structurally, the lack of predefined standards in the prompts led to some overly broad responses from ChatGPT. Without guidance on what constitutes "useful" evidence, ChatGPT turned out to generate generic paraphrasing, which Li found redundant and ineffective. This structural issue could pose barriers to use ChatGPT as a writing partner as Li showed dissatisfaction towards ChatGPT's response. To address this, our revised intervention incorporated specific formats and standards into the prompts: asking ChatGPT to

provide evidence using concrete examples of real-life scenarios.

Discussions

Following Yang and Carless (2013)'s feedback triangle, we have iteratively refined the ChatGPT-mediated dialogic feedback design, which was initially inadequate but later perceived as effective by participants. The above findings lead us to propose that to foster effective dialogues, ChatGPT-mediated feedback activities should involve (1) providing discipline-specific guidance to foster disciplinary knowledge, (2) promoting students' self-evaluation of their work before turning to ChatGPT, (3) incorporating graduated mediation with a scoring element to enhance cue-consciousness, (4) adapting the roles of ChatGPT to different writing stages, (5) nurturing a critical attitude towards ChatGPT's responses, (6) mobilising learners' full tool repertoire, and (7) developing feedback structures based on specific formats and standards. Each of the seven proposed suggestions maps onto our three-dimension and seven-element feedback triangle in Figure 5: suggestions 1 – 3 target the cognitive dimension, 4 – 5 the social-affective, and 6 – 7 the structural. This section will discuss each dimension and outline the corresponding design elements and suggestions.

The cognitive dimension refers to the feedback content to support addressing disciplinary problems and self-regulated learning (Yang & Carless, 2013). One disciplinary problem in argumentative writing is how to orchestrate diverse structural components into one coherent work (Aull & Ross, 2020). This calls for specific guidance in IELTS writing structures to foster disciplinary learning, especially given ChatGPT's reliance on generalised rather than domain-specific knowledge (Hu et al., 2024). Thus, incorporating discipline-specific guidance into feedback content can help foster disciplinary knowledge, identified as the first cognitive attribute that the ChatGPT-mediated feedback practices target at. More importantly, productive feedback practices should go beyond the development of disciplinary knowledge to nurture students' self-regulation: the planning, monitoring, and evaluation of learning, and the adaptation of learning strategies to task demands and progress (Pekrun et al. 2002). This study then spotlights two attributes of self-regulative learners: self-evaluation and cue-consciousness, engaging learners in a cycle from their self-evaluation of their writing to ChatGPT's graduated mediation. The cycle starts with learners' self-evaluation, where learners assess their strengths and weaknesses according to task criteria, reducing reliance on the feedback provider (Nicol & MacFarlane-Dick, 2006), specifically ChatGPT in this study. After the self-evaluation, graduated mediation through ChatGPT should be provided for writers, namely, implicit to explicit support. Specific scores and corresponding explanations should be incorporated into the feedback, which were explicit cues about disciplinary priorities and criteria and could promote learners' cue-consciousness. Therefore, such a cycle from learners' self-evaluating their work to ChatGPT providing graduated mediation is vital for learners to make constructive use of feedback and become self-regulative (Yang & Carless, 2013). In sum, the analysis of the cognitive dimension leads us to propose that discipline-specific guidance should be given to students to help them

address disciplinary problems and foster disciplinary knowledge. Designing a continuous cycle that guides students from self-evaluating their work to eliciting graduated mediation with scoring from ChatGPT can enhance their self-evaluation capacity and cue-consciousness.

The social-affective dimension highlights feedback as a social practice where the relationship between feedback providers and receivers can shape the learning experience (Yang & Carless, 2013). This research demonstrates how ChatGPT's roles can be adapted across writing stages, and how a critical attitude towards ChatGPT can be nurtured. Specifically, GenAI-driven chatbots have predominantly been employed to provide explicit corrective feedback (Mizumoto & Eguchi, 2023; Poole & Coss, 2024). This may fail to recognise the uniqueness of GenAI-driven chatbots who can go beyond a simple tool function to play multiple roles as a full participant in learning (Sharples, 2023). In this research, through different prompts, ChatGPT has been positioned differently at different writing stages, showcasing the adaptability of GenAI in providing support for different needs (Sharples, 2023). For instance, ChatGPT was used as a brainstorming peer to generate alternative perspectives before writing and as a professional assessor to evaluate one's writing against the criteria after writing. GenAI chatbots can thus move beyond simple prompts and responses to become a full participant in writing process (Sharples, 2023). Notably, the feedback activities should be designed to foster students' critical engagement with ChatGPT and their sense of agency. Unlike a trusting relationship in Yang and Careless (2013)'s design, an untrusting relationship with ChatGPT and critical attitudes have been outlined in our study. Some research has found ChatGPT to be accurate in assessing linguistic errors, suggesting its potential as a tool for written corrective feedback (Jiang et al., 2023; Long, 2024; Gozali et al., 2024), though this is not always the case in our study. For instance, ChatGPT once suggested rephrasing a grammatically correct sentence, which altered the writer's intended meaning, possibly due to ChatGPT's generative nature rather than error detection. This aligns with concerns in recent research about GenAI's tendency to produce false and misleading content, which learners may struggle to recognise (Abdelhalim, 2024; Yuan et al., 2024). It is thus essential for learners to critically evaluate GenAI-generated feedback regarding its accuracy, relevance, and authenticity (Chiu, 2023; Jeon & Lee, 2023; Walter, 2024). Overall, from a pedagogical design perspective, practitioners should design ChatGPT-mediated feedback activities that assign context-specific roles to ChatGPT and foster learners' critical attitudes towards it, thereby fostering learner agency and effective uptake.

The structural dimension focuses on how the feedback is organised, and resources are deployed to overcome structural challenges (Yang & Carless, 2013). Traditional challenges such as large classes and heavy assessment workloads (Beaumont et al., 2011) can be addressed through our design, which leverages a chatbot to provide immediate, personalised feedback throughout the writing process. However, new structural challenges arose from our research process, concerning the functionalities of ChatGPT. The first one concerns the mobilisation of tool resources since ChatGPT shares some overlapping functions with many other tools. Some

researchers have positioned ChatGPT as an alternative to current tools (e.g. Zhang et al., 2023), while our participants viewed ChatGPT as part of a holistic tool repertoire to co-afford writing. Students can thus employ both GenAI and traditional tools as a whole tool repertoire to complete writing (Xiao & Zhi, 2023). The second challenge in this study was caused by GenAI's inherent limitations to produce overly vague answers and the prompt designs. Limited prompts in our research led to writers' dissatisfaction, while our iteratively improved ones can guide the feedback process. This structural design element on prompting aims to preset expectations and criteria for ChatGPT's feedback, which can help plan and coordinate the feedback activities and establish a common ground for dialogues (Hadwin et al., 2017; Erkan et al., 2021). Altogether, our analysis highlights two design features of ChatGPT-mediated dialogic feedback: integrating GenAI and other tools into a holistic repertoire for learners to use, and structuring prompts with specific formats and standards to mitigate the limitations of GenAI.

To sum up, grounded in the notion of dialogic feedback, our framework outlines the necessary design elements and dimensions of an optimal ChatGPT-mediated dialogic feedback activity. The cognitive dimension is at the top of this triangle, suggesting that the feedback content is the most central dimension. Priority should thus be paid to the corresponding design elements (i.e. disciplinary knowledge, self-evaluation, and cue-conscious). The social-affective and structural dimensions serve as foundations to support the cognitive dimension, indicating that if the other elements are not addressed, the central three cognitive attributes in student learning cannot be developed. Seven design principles are derived from the three dimensions of feedback practice as aforementioned and can be used in combination to design ChatGPT-mediated dialogic feedback.

Conclusion

Through a DBR approach, we developed an intervention to guide the design of ChatGPT-mediated dialogic feedback activities. Seven design elements were derived from and adjusted according as participants' practices and mapped onto a dialogic feedback triangle. From this triangle, we drew out seven practical implications for future educators to design and practice ChatGPT-mediated dialogic feedback. Theoretically, the proposed three-dimension and seven-element feedback triangle enriches and extends Yang and Carless's (2013) feedback framework to the context of GenAI chatbots, offering a structured way to theorise ChatGPT-mediated feedback processes. Echoing a call for the re-conceptualisation of feedback (Yang & Carless, 2013), we proposed a framework capturing the cognitive, social-affective, and structural nature of GenAI-mediated feedback activities. Methodologically, this research demonstrates the potential of a DBR approach in iteratively refining feedback designs and improving their effectiveness. Our dual positionality as both researchers and co-designers of the intervention offers insights into participatory design in DBR, showing the importance of collaboration between researchers, educators, and learners in developing interventions. Practically, the proposed

framework offers a simple and flexible yet comprehensive and actionable guide for educators and developers to design ChatGPT-mediated dialogic feedback. It highlights key design elements to foster effective dialogues in pedagogical practices and can inform the technological development of GenAI writing tools.

However, this research is not without limitations. This research was purely qualitative to reveal the development process of the feedback design, while its effectiveness could be examined through a mixed-method approach in future. Meanwhile, the sample size was limited, and participants' high proficiency levels and familiarity with digital tools might have introduced potential biases. The findings may not fully apply to less proficient learners or different writing tasks. Future research can include a larger and more diverse group with varying proficiency levels to enhance the generalisability. Additionally, the three-dimensional dialogic feedback triangle can be applied in combination with other models (e.g. a three-phase collaborative dialogic feedback framework from Erkan et al. (2021) to integrate ChatGPT-mediated dialogic feedback into curriculum. Future researchers can combine other models with our feedback triangle to structure the GenAI-mediated dialogic feedback activities.

References

Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a Foreign Language undergraduates' perceptions and practices across varied metacognitive awareness levels. Journal of Computer Assisted Learning, 40(3), 1261 – 1275. https://doi.org/10.1111/jcal.12948

Ajjawi, R., & Boud, D. (2017). Researching Feedback Dialogue: An Interactional Analysis Approach. *Assessment &Evaluation in Higher Education 42*(2), 252–265. doi:10.1080/02602938.2015.1102863.

Akker, J. J. H. van den. (2006). *Educational design research.* Routledge.

Alejandro, A. (2020). Reflexive discourse analysis: A methodology for the practice of reflexivity. European Journal of International Relations, 27(1), 150-174. https://doi.org/10.1177/1354066120969789 (Original work published 2021)

Alkaissi, H., & McFarlane, S. I. (2023). Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus, 15*(2), Article e35179. https://doi.org/10.7759/cureus.35179

Andrade, C., & Roshay, A. (2024). Using Google Docs for collaborative writing feedback with international students. The CATESOL Journal, 34(1).

Aull, L. L., & Ross, V. (2020). From cow paths to conversation: Rethinking the argumentative essay. *Pedagogy: Critical Approaches to Teaching Literature, Language, Composition, and Culture, 20*(1), 21–34. https://doi.org/10.1215/15314200-7878975

Bakker, A. (2018). *Design Research in Education: A Practical Guide for Early Career Researchers* (1st ed.). Routledge. https://doi.org/10.4324/9780203701010

Barrot. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57,* 100745. https://doi.org/10.1016/j.asw.2023.100745

Beaumont, C., O' Doherty, M., & Shannon, L. (2011). Reconceptualising assessment feedback: A key to improving student learning? Studies in Higher Education, 36(6), 671‑687. https://doi.org/10.1080/03075071003731135

Bloxham, S., & Campbell, L. (2010). Generating dialogue in assessment feedback: Exploring the use of interactive cover sheets. *Assessment & Evaluation in Higher Education, 35*(3), 291–300. https://doi.org/10.1080/02602931003650045

Boekaerts, M. (2010). The crucial role of motivation and emotion in classroom learning. In The nature of learning: Using research to inspire practice, ed. H. Dumont, D. Istance and F. Benavides, 91 111. Paris: OECD.

Braun, & Clarke, V. (2022). Thematic analysis: A practical guide.

Carless, D. (2016). Feedback as dialogue. In M. Peters (Ed.), *Encyclopedia of educational philosophy and theory* (pp. 1–6). Springer. https://doi.org/10.1007/978-981-287-532-7

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). *Developing sustainable feedback practices. Studies in Higher Education, 36*(4), 395-407.

Chiu, T. K. F. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: A case of ChatGPT and Midjourney. Interactive Learning Environments. Advance online publication. https://doi.org/ 10.1080/10494820.2023.2253861

Dowden, T., Pittaway, S., Yost, H., & McCarthy, R. (2013). Students' perceptions of written feedback in teacher education: Ideally feedback is a continuing two-way communication that encourages progress. Assessment & Evaluation in Higher Education, 38(3), 349 ‑ 362. https://doi.org/10.1080/02602938.2011.632676

Erkan Er, Yannis Dimitriadis, & Dragan Gašević. (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

Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. Journal of Second Language Writing, 22(3), 307–329. https://doi.org/10.1016/j.jslw.2012.09.009

Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. Computers and Education: Artificial Intelligence, 6, Article 100250. https://doi.org/10.1016/j.caeai.2024.100250

Gozali, I., Wijaya, A. R. T., Lie, A., Cahyono, B. Y., & Suryati, N. (2024). Leveraging the potential of ChatGPT as an automated writing evaluation (AWE) tool: Students' feedback literacy development and AWE tools integration framework. The JALT CALL Journal, 20(1), 1 – 22. https://doi.org/10.29140/jaltcall.v20n1.1200

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies.* https://doi.org/10.1007/s10639-023-12146-0

Hadwin, A. F., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In D. H. Schunk & J. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 201–215). Routledge.

Hu, W., Tian, J., & Li, Y. (2024). Enhancing student engagement in online collaborative writing through a generative AI-based conversational agent. *The Internet and Higher Education.* https://doi.org/10.1016/j.iheduc.2024.100979

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Hui, L., Ippolito, K., Sarsfield, M., & Charalambous, M. (2024). Using a self-reflective ePortfolio and feedback dialogue to understand and address problematic feedback expectations. Assessment and Evaluation in Higher Education, 49(3), 334 – 347. https://doi.org/10.1080/02602938.2023.2232960

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. Education and Information Technologies, 28, 15873 – 15892. https://doi.org/10.1007/s10639-023-11834-1

Jiang, Z., Xu, Z., Pan, Z., He, J., & Xie, K. (2023). Exploring the role of artificial intelligence in facilitating assessment of writing performance in second language learning. Languages (Basel, Switzerland), 8(4), 247. DOI: 10.3390/languages8040247

Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. Interactive Learning Environments. Advance online publication. https://doi.org/10.1080/10494820.2025.2498537

Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. Interactive Learning Environments, 1 – 25. https://doi-org.eproxy.lib.hku.hk/10.1080/10494820.2025.2498537

Long, H. S. (2024). Exploring the use of ChatGPT as a tool for written corrective feedback in an EFL classroom. The Journal of Asia TEFL, 21(2), 397 – 412. https://doi.org/10.18823/asiatefl.2024.21.2.8.397

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4373111

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

OpenAI. (2024, January 10). Usage policies. OpenAI. https://openai.com/policies/usage-policies

Pekrun, R., Goetz, T., & Titz, W. (2002). Academic Emotions in Students' Self Regulated Learning and Achievement: A Program of Quantitative and Qualitative Research. Educational Psychologist, 37, 91-105. https://doi.org/10.1207/S15326985EP3702_4

Poole, F. J., & Coss, M. D. (2024). Can ChatGPT Reliably and Accurately Apply a Rubric to L2 Writing Assessments? The Devil is in the Prompt (s). *Journal of Technology & Chinese Language Teaching, 15*(1). https://doi.org/10.35542/osf.io/3r2zb

Sharples, M. (2023). Towards social generative AI for education: Theory, practices and ethics. *Learning: Research and Practice, 9*(2), 159–167. https://doi.org/10.1080/23735082.2023.2261131

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57,* 100752. https://doi.org/10.1016/j.asw.2023.100752

Sutton, P., & Gill, W. (2010). Engaging feedback: Meaning, identity and power. *Practitioner Research in Higher Education, 41*(1), 3–13.

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. Computers and Education: Artificial Intelligence, 7, 100270. https://doi.org/10.1016/j.caeai.2024.100270

The Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. Educational Researcher, 32(1), 5-8. https://doi.org/10.3102/0013189X032001005

Van Hoe, A., Wiebe, J., Slotta, J., Rotsaert, T., & Schellens, T. (2024). Designing Dialogic Peer Feedback in Collaborative Learning: The Role of Thinq Tank. Education Sciences, 14(11), 1231. https://doi.org/10.3390/educsci14111231

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

Wegerif, R. (2013). Dialogic: Education for the internet age. Routledge.

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

Wood, J. (2023). Enabling feedback seeking, agency and uptake through dialogic screencast feedback. Assessment & Evaluation in Higher Education, 48(4), 464 – 484. https://doi.org/10.1080/02602938.2022.2089973

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages, 8*(3), 212. https://doi.org/10.3390/languages8030212

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*(11), 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Yang, M. & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes.*Teaching in Higher Education.18*(3), 285-297.

Yuan, Y., Li, H., & Sawaengdist, A. (2024). The impact of ChatGPT on learners in English academic writing: Opportunities and challenges in education. Language Learning in Higher Education, 14(1), 41 – 56. DOI: 10.1515/cercles-2023-0006

Zhang & Liu (Accepted, Forthcoming). Tame the Beast of ChatGPT: Developing Design Principles to Strategically Integrate ChatGPT into EFL Writing through an Activity Theory Perspective. *Computer Assisted Language Learning*

Zhang, R., Zou, D., & Cheng, G. (2023). Chatbot-based learning of logical fallacies in EFL writing: Perceived effectiveness in improving target knowledge and learner motivation. *Interactive Learning Environments.* https://doi.org/10.1080/10494820.2023.2220374

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. Higher Education Research & Development, 37(4), 883-897.

Appendix A. Evolution of the Intervention

The following tables present the revisions made to our interventions in Iteration 1 and 2, with no revisions in Iteration 3. Corresponding design elements and recommendations are attached.

Revisions to the activity design with screenshots of the revised intervention	Corresponding design elements and recommendations
Iteration 2: Some guidance on writing IELTS essays was added to guide writers during content drafting stage. ➢ The following structure comes from a famous and experienced IELTS teacher and previous examiner Simon (Please note that this is for reference, and you can structure your writing specifically regarding the question) 1. **Phase 1: Writing an introduction paragraph** (paraphrasing the task and introducing your main claim) **2. Phase 2: Writing body paragraphs based on your previous sub-claim - evidence - reasoning structure** Body paragraph 1: Body paragraph 2: **3. Phase 3: Writing a conclusion**	**Disciplinary knowledge:** providing discipline-specific guidance to foster disciplinary knowledge

<table>
<tr>
<td>Iteration 1:
During outline generation, notes were added to encourage writers to first brainstorm and document their ideas before consulting ChatGPT.

1.(Optional) Brainstorm with ChatGPT by asking about the explanation of key concepts or the writing topic.
“Provide a short definition of XXX (the keywords in the writing task) ”
“What is the nature of XXX? Give me a short answer.”
Or: You could send the writing question to ChatGPT after you have some initial ideas

2.Write an initial outline for the task first on your own and follow a CER (claim-evidence-reasoning) framework

3.Compare and evaluate ChatGPT-generated outline with yours

During content drafting, an additional note emphasised the importance of maintaining agency and avoiding direct copying of ChatGPT’s output.

Stage 2-Content Drafting
➢ During the 1-3 phases, you can ask questions to ChatGPT but cannot ask it to write paragraphs for you.
➢ You should not totally copy or paste contents from ChatGPT or other resources.</td>
<td>Self-evaluation: promoting students’ self-evaluation of their work before turning to ChatGPT</td>
</tr>
<tr>
<td>Iteration 1:
Graduated mediation - from implicit to explicit feedback - and a scoring element were incorporated

Stage 3-Reflection and Proofreading
During this stage, you are advised to first gain comprehensive and general feedback on the four writing criteria, with specific scoring from ChatGPT. Please note that the scoring from ChatGPT cannot predict your actual scores in exams, but serves as a reflective tool for you to improve your writing. After gaining an overall evaluation, you can then choose to focus on some specific dimensions, ask for focused feedback, and revise your writing.

1. Phase 1: General Evaluation

2. Phase 2: Specific Evaluation and Proofreading
You can ask ChatGPT to specifically provide feedback on any of the four dimensions and then revise your writing.</td>
<td>Cue-consciousness: incorporating graduated mediation with a scoring element to enhance cue-consciousness</td>
</tr>
</table>

Iteration 2: More suggested prompts for focused feedback were added. **2. Phase 2: Specific Evaluation and Proofreading** You can ask ChatGPT to specifically provide feedback on any of the four dimensions and then revise your writing. For **Task Response and Cohesion and Coherence**, you are advised to ask follow-up questions for clarification, explanation, and correction directly according to ChatGPT's evaluation. Here are some suggested ways: *"Point out how..." "Give me concrete examples of ..."* ***"Where can I add cohesive devices in my writing?" "How can I strengthen my reasoning in XXX?"***	
Iteration 1: ChatGPT was prompted differently at different writing stages (as a brainstorming peer before writing and professional assessor after writing). (1) A step to "brainstorm with ChatGPT" was added with suggested prompts 1.(Optional) Brainstorm with ChatGPT by asking about the explanation of key concepts or the writing topic. "Provide a short definition of XXX (the keywords in the writing task) " "What is the nature of XXX? Give me a short answer." Or: You could send the writing question to ChatGPT after you have some initial ideas (2) A step to "compare and evaluate ChatGPT-generated outline with yours" was added with a prompt asking ChatGPT to independently generate an outline instead of providing feedback on writers' outlines	**Adaptive Roles of ChatGPT at Different Writing Stages:** adapting the roles of ChatGPT to different writing stages

3.Compare and evaluate ChatGPT-generated outline with yours "Act as an excellent argumentation outliner who follows the following structure and criteria to answer an IELTS Task 2: IELTS task 2: (Please insert the question here) Structure: Main claim: (Please insert your main claim here) Sub claim 1: Evidence 1: Reasoning 1: Sub claim 2: Evidence 2: Reasoning 2: Criteria: the sub-claims are closely connected and progressed logically to support the main claim the evidence can support the sub-claims the reasoning provides a justification of why the evidence support the sub-claim" (3) ChatGPT was prompted both for general evaluation and focused feedback at proofreading stage **Stage 3-Reflection and Proofreading** During this stage, you are advised to first gain comprehensive and general feedback on the four writing criteria, with specific scoring from ChatGPT. Please note that the scoring from ChatGPT cannot predict your actual scores in exams, but serves as a reflective tool for you to improve your writing. After gaining an overall evaluation, you can then choose to focus on some specific dimensions, ask for focused feedback, and revise your writing. **1. Phase 1: General Evaluation** **2. Phase 2: Specific Evaluation and Proofreading** You can ask ChatGPT to specifically provide feedback on any of the four dimensions and then revise your writing.	
Iteration 1: Some guidance on ChatGPT's limitations was incorporated at every stage: writers are advised that ChatGPT may provide vague or inconsistent comments and are encouraged to run their writing through ChatGPT multiple times and cross-check the consistency of feedback. At each stage, ChatGPT may provide vague comments, so you are highly recommended to ask follow-up questions to ChatGPT. ChatGPT may also provide inconsistent answers, so you may run the writing through ChatGPT multiple times and check the consistency of the feedback.	**A Critical Attitude towards ChatGPT:** nurturing a critical attitude towards ChatGPT's responses
Iteration 1: A suggestion on using e-dictionaries for authentic word-searching was added. ➢ You are encouraged to use e-dictionary to seek vocabulary and its authentic use, and any other available resources are encouraged to be used.	**Mobilisation of Learners' Full Tool Repertoire:** mobilising learners'

A suggestion on using search engines for information check was added. ➢ You are encouraged to use search engines (e.g. Google, Baidu) to check and enrich the evidence part of your outline	full tool repertoire
Iteration 2: The criteria for ChatGPT-generated examples were made clearer as: "The evidence can involve real-world examples to support the sub-claims," 5. Phase 5: Compare and evaluate ChatGPT-generated outline with yours "Act as an excellent argumentation outliner who follows the following structure and criteria to answer an IELTS Task 2: IELTS task 2: (Please insert the question here) Structure: Main claim: (Please insert your main claim here) Sub claim 1: Evidence 1: Reasoning 1: Sub claim 2: Evidence 2: Reasoning 2: Criteria: The sub-claims are closely connected and progressed logically to support the main claim; **The evidence can involve real-world examples or expert quotes to support the sub-claims** the reasoning provides a justification of why the evidence support the sub-claim"	**Feedback Structures based on Specific Formats and Standards:** developing feedback structures based on specific formats and standards