

Chinese Undergraduate L2 Writers' Metacognitive Regulation When Using Teacher and AI Feedback: A Qualitative Study

Shuying Diao ^{1,*}

¹ College of Education, University of Southampton, Southampton, Hampshire, SO17 1BJ, United Kingdom

* Correspondence author: sydiaoedu@phc.ac.cn

Abstract: Generative artificial intelligence is increasingly used alongside teacher feedback in second-language writing, yet limited attention has been paid to how learners regulate their writing when engaging with these two feedback sources. This qualitative study explored the reported metacognitive regulation of five Chinese undergraduate non-English majors who had experience using both teacher feedback and generative AI tools in L2 writing. Data were collected through semi-structured interviews and analysed using reflexive thematic analysis. The findings indicated considerable variation in participants' planning, monitoring, checking and revision practices. Some participants described systematic regulation involving outlining, paragraph-level monitoring and revisions extending beyond explicitly marked problems, whereas others mainly corrected errors directly identified by teachers or AI tools. Most participants regarded teacher feedback as more authoritative for content, organisation, genre and assessment expectations, while AI feedback was valued primarily for its immediacy and support with grammar, vocabulary and sentence-level expression. Participants who reported clearer goals and more active monitoring also perceived greater benefits from feedback, although the study does not establish objectively measured improvement in writing ability. The findings highlight the importance of feedback literacy and learner agency in human-AI feedback environments. Pedagogically, students may benefit from explicit guidance on comparing, evaluating and selectively applying feedback from different sources.

Keywords: L2 writing; metacognitive regulation; teacher feedback; generative AI; feedback literacy; qualitative research

1. Introduction

In many non-English contexts, second language (L2) writing has become an important academic skill for university students. It is closely related to students' academic performance, future career opportunities, and participation in international academic and professional communication. In China, this connection is particularly evident: the national higher education policy emphasizes undergraduates' English proficiency and academic literacy and identifies writing as a key learning outcome of college English courses [1]. At the same time, L2 writing remains challenging for many non-native English speakers. This is not just a matter of possessing sufficient linguistic knowledge; learners need to deal with language forms, adhere to discipline and rhetorical practices, and manage the high-order thinking process at the same time [2].

Writing well involves much more than knowing the language. It also requires writers to regulate and monitor their own thinking during the writing process. L2 writing has often been characterised as a complex cognitive activity in which writers set goals, plan what they want to say, monitor how the text is progressing, assess what has been produced and then revise as necessary [3]. These activities are central to metacognitive regulation. Metacognitive regulation includes setting appropriate goals,



engaging in strategic planning, monitoring one's progress, assessing the quality of the emerging text, and revising on the basis of both external feedback and self-evaluation [4]. During the writing process, monitoring helps L2 writers track their progress and maintain an awareness of whether their text meets task demands. Together, these metacognitive processes enable writers to identify problems, allocate cognitive resources, and apply corrective strategies, thereby potentially supporting both immediate text revision and longer-term writing development [5].

Feedback has long been regarded as a key tool for improving writing and fostering the development of learners' metacognitive skills. Traditionally, teachers have been the primary external source of guidance, providing subject-matter knowledge, explanatory feedback, evaluative judgments, and motivational support. In recent years, feedback research has increasingly shifted toward a learner-centered perspective, emphasizing how students actively interpret, evaluate, and act upon feedback rather than passively accepting it. At the same time, with the rapid development of artificial intelligence and writing assistance technologies, a variety of digital and AI-assisted tools have emerged in the classroom. Examples include automatic grammar checkers, rule-based or statistical automatic writing evaluation (AWE) systems, and recently, generative feedback tools (such as language models). These tools can identify grammatical errors and suggest revisions more quickly and consistently. Research by Chen and Cheng (2008) indicates that when AWE is used in conjunction with teacher guidance, this blended feedback in classroom writing not only helps students improve text quality through multiple revisions but is also more readily accepted by students as a useful learning resource [6]. Consequently, the collaborative use of teachers and AI feedback is becoming increasingly common in higher education, providing students with more timely and diverse sources of feedback; however, the specific outcomes still depend on how the feedback is integrated and on instructional design.

However, existing research on human-AI collaborative feedback in second-language writing has primarily focused on outcome-orientated measures (e.g., improved scores and reduced errors), while the learners' internal cognitive regulatory experiences have received scant attention. In this study, AI feedback refers to responses generated by tools such as ChatGPT and Doubao on students' drafts, including comments or suggestions concerning grammar, vocabulary, organisation, and expression. Human-AI collaborative feedback refers to students' use, comparison, and integration of teacher feedback and AI-generated feedback during drafting or revision; it does not refer to a formally integrated teacher-AI feedback system. Existing research on L2 writing feedback has provided substantial evidence that both teacher feedback and AI feedback can, to some extent, enhance writing quality. However, most of this work still focuses primarily on product-oriented outcomes such as scores, accuracy, and holistic text quality, while overlooking the importance of learners' cognitive and metacognitive processes during writing. In particular, little is known about how the combination of teacher and AI feedback shapes students' planning, monitoring, and revision behaviours. Addressing this gap is essential, because changes in metacognitive regulation are likely to underpin continuous improvements in writing ability and more autonomous in writing practice.

Therefore, the present study investigates how Chinese undergraduate L2 writers report planning, monitoring, checking, and revising their texts when teacher feedback and AI-generated feedback are both available. It also examines how participants perceive the distinct roles of the two feedback sources and how they describe the relationship between their regulatory practices and their writing processes.

Drawing on the concept of metacognitive regulation and set against the backdrop of English writing among Chinese undergraduates, this study investigates how students describe their metacognitive regulation when using teacher and AI feedback. Specifically, this study examines the following three research questions:

- (1) How do Chinese undergraduate L2 writers describe their planning, monitoring, checking, and revising practices when completing English writing tasks?
- (2) How do students perceive the relationship between metacognitive regulation and their L2 writing processes and performance?
- (3) How do students perceive, compare, and respond to teacher feedback and AI-generated feedback during L2 writing and revision?

The contribution of this study lies in providing an exploratory qualitative account of how a small group of Chinese undergraduate L2 writers interpret and use teacher and AI-generated feedback during planning, monitoring, checking, and revision. Rather than testing the effectiveness of human-AI collaborative feedback or establishing a causal relationship between metacognitive regulation and writing achievement, the study identifies reported patterns, perceived benefits, and challenges within this feedback context.

2. Literature Review

2.1. *Teacher Feedback in L2 writing*

Feedback on student writing has long been one of the most extensively studied topics in L2 writing. A large body of research indicates that teacher feedback can support students' revision and writing development when learners have meaningful opportunities to interpret and act on it. Classroom-based research has shown that teacher written feedback may contribute to improvements in accuracy, complexity, and overall text quality, particularly when students are required and supported to revise their drafts [3]. From a broader educational perspective, feedback is often described as information that helps to narrow the gap between learners' current performance and intended learning goals. It can shed light on specific aspects of students' writing, highlight assessment criteria more clearly and prompt them to reflect on or alter their approaches [7]. This understanding shifts attention from feedback as information delivered by teachers to feedback as a process that requires learners to interpret comments and translate them into revision decisions.

Much work on L2 writing has concerned corrective feedback and its impact on accuracy. One well-known study by Bitchener, Young and Cameron (2005), for instance, compared direct correction, metalinguistic comments and more implicit feedback in ESL writing classes [8]. They found that explicit written correction was associated with improved use of English articles in revised drafts and subsequent writing tasks. Other research has suggested that more indirect forms of feedback, such as indicating where an error is without supplying the correct form, can also be helpful, particularly for learners who are willing and able to engage in problem-solving and work out the answer for themselves. Such feedback may encourage learners to diagnose problems and consider alternative forms independently. However, its effectiveness varies according to error type, learner proficiency, prior knowledge, and the degree of engagement with the feedback. Moreover, much of the strongest evidence concerns linguistic accuracy rather than higher-order aspects of writing.

Some studies also examined teachers' broader comments on aspects such as content, organisation and argumentation. For example, Hyland and Hyland (2006) reported that students generally value teacher comments that explain the reasons behind suggestions and that link feedback to assessment criteria [9]. Research conducted in Asian English as a Foreign Language (EFL) contexts, including Hong Kong and mainland China, indicates that learners typically perceive teacher feedback as authoritative and exam-orientated, and that they tend to prioritise comments directly related to their grades [10]. At the same time, students may find feedback overly extensive, insufficiently organised, or difficult to translate into specific revision actions. These findings suggest that the usefulness of teacher feedback depends not only on its content but also on its clarity, prioritisation, and actionability.

Alongside teacher feedback, automated writing evaluation has become an important strand of L2 writing research. Early systems such as e-rater were primarily designed to generate automated scores and identify selected textual features in large-scale assessment contexts. Evaluations reported correlations between automated and human scores, suggesting that such systems could approximate some dimensions of overall writing performance under particular assessment conditions [11]. In classroom contexts, automated tools have also been used to provide rapid feedback on grammar, vocabulary, and basic aspects of coherence. For example, Li, Link and Hegelheimer (2015) found that ESL students employed automated writing assessment feedback to fix many surface errors and raise the scores of their drafts [12]. Students valued the speed and volume of feedback because it allowed them to conduct multiple rounds of revision within a relatively short period.

Despite these advantages, automated feedback has important limitations. Earlier automated systems often struggled with context-dependent language problems and higher-order rhetorical features [13]. Qualitative work suggests that some learners accept the suggestions made by automation almost uncritically, even when the suggestions are inappropriate, because they perceive the system as neutral and reliable. Others just dismiss feedback they don't understand, particularly if the explanations are brief or very rule-based. There are also concerns that automated feedback promotes a narrow focus on sentence-level issues and may detract from content development, audience awareness and critical thinking [12]. Overall, these findings suggest that automated tools may be most useful when embedded within a broader pedagogical framework rather than used as stand-alone sources of writing guidance.

2.2. *AI-assisted feedback and human- AI collaborative feedback*

Teacher feedback commonly addresses both linguistic form and higher-order concerns, including content, organisation, argumentation, genre, and task fulfilment [14]. By comparison, automated and AI-assisted feedback has traditionally been used more extensively for rapid language-level checking, although newer generative AI tools can also produce comments on organisation, content, and style.

Classroom studies indicate that students can improve the quality of their drafts with the use of automated feedback as part of their regular instruction and are generally receptive to this type of support as convenient and helpful [12]. Nevertheless, learner engagement with AI feedback is uneven. Some learners focus mainly on surface-level correction, some accept suggestions without sufficient evaluation, and others ignore or reject comments that appear vague, inappropriate, or difficult to interpret.

Recent generative AI tools differ from earlier automated writing evaluation systems because they can generate open-ended explanations, engage in iterative dialogue, and respond to user prompts. Classroom studies and mixed-methods research indicate that, under teacher guidance, learners use generative AI to brainstorm, organize, and revise their work multiple times [15]. However, generative responses may also be inaccurate, overly general, stylistically inappropriate, or insufficiently aligned with specific course requirements. Scholars warn against overreliance on generative answers unless students engage in critical evaluation and possess GenAI literacy [16]. Therefore, learners need to evaluate rather than automatically accept AI-generated suggestions. They should incorporate human feedback, prompt refinement, and verification practices into their revision process. However, generative responses may also be inaccurate, overly general, stylistically inappropriate, or insufficiently aligned with specific course requirements. Therefore, scholars emphasize that students must move away from passively accepting AI recommendations. Instead, they should develop critical AI literacy in order to evaluate and verify machine-generated content [15].

Research directly examining the combined use of teacher and AI feedback in L2 writing remains comparatively limited. Existing work suggests that learners may value automated feedback for its immediacy and assistance with identifiable language problems, while relying more heavily on teachers for contextualised explanations, higher-order guidance, and clarification of assessment expectations [17]. Some students say they don't know what to do when the two sources disagree, and others just go with the easiest option, without weighing up the alternatives. These patterns indicate that simply providing multiple feedback sources does not ensure productive engagement. Learners may also require guidance in comparing, evaluating, and integrating feedback from different sources.

Across these strands of research there has been a growing interest in how learners actually use feedback and what is now often called "feedback literacy". From this perspective, feedback does not produce learning automatically; its value depends partly on learners' capacity to understand, evaluate, and act on it. For example, Carless and Boud (2018) argue that effective feedback depends not just on what teachers or systems say, but also on learners' ability to interpret comments, judge their value and translate them into concrete action [18]. Research on student participation indicates that some learners will carefully read the feedback, compare opinions from different sources, and formulate targeted revision plans. In contrast, some learners will skim the feedback, make superficial corrections, and then proceed with their learning without reflecting on it. Beliefs about feedback, perceptions of fairness, and past experiences will all affect these patterns, leading to varying degrees of engagement and effectiveness in utilising feedback for learning improvement. Feedback literacy therefore provides a useful basis for examining why learners engage differently with the same or similar feedback and how they negotiate the respective authority of teachers and AI systems.

2.3. Metacognitive regulation

Although technology-assisted feedback has become increasingly widespread, the literature examining feedback from the perspective of metacognitive regulation remains relatively scarce. Metacognitive regulation is generally understood as a series of processes in which learners set goals, plan, monitor, and evaluate their own learning, and adjust their strategies based on the results of that evaluation [4]. In second-language writing, these processes may include: clarifying task objectives, planning content and structure, monitoring ideas and language during the writing process, evaluating whether the text meets communicative and assessment requirements, and making conscious revisions based on self-assessment or external feedback.

Metacognitive regulation is shaped by multiple learner- and context-related factors, including self-efficacy, prior strategy use, feedback literacy, task demands, and classroom conditions. Learners with stronger self-efficacy and a more developed repertoire of writing strategies may be more likely to set specific goals, monitor their progress, and use feedback to adjust subsequent writing practices. In L2 writing, self-regulated learning strategies have been conceptualised as multidimensional practices involving cognitive, metacognitive, social-behavioural, and motivational regulation, highlighting the importance of planning, monitoring, and strategic adjustment throughout the writing process [19]. However, much empirical work on feedback still defines success primarily in product terms, including higher scores, fewer errors, and improved holistic ratings. Comparatively less attention has been paid to the processes through which learners incorporate feedback into planning, monitoring, evaluation, and

revision. In particular, limited qualitative evidence is available on how learners describe changes in their planning, monitoring, checking, and revision when teacher and AI feedback coexist. Existing studies often examine teacher feedback, automated feedback, or self-regulated writing separately, leaving the interaction among feedback source, learner evaluation, and metacognitive regulation insufficiently understood. This gap is especially relevant in relation to Chinese non-English-major undergraduates, whose writing experiences may be shaped by assessment expectations, varied English proficiency, and differing levels of familiarity with generative AI tools.

In summary, existing research suggests that teacher and AI-assisted feedback may offer different but potentially complementary forms of support for L2 writing. Teacher feedback is commonly valued for its contextualised guidance on content, organisation, genre, and assessment expectations, whereas automated and AI-generated feedback is often valued for its immediacy and assistance with language-level revision. Empirical study reports the positive impact on accuracy, overall text quality and revision behavior, and points out some factors that seem to shape these influences, including the proficiency of learners, their feedback literacy, and the way of systematic design and integration into teaching [12]. However, the coexistence of teacher and AI feedback does not necessarily result in productive collaboration, because learners may interpret, prioritise, accept, or reject the two sources in different ways.

Nevertheless, an important gap remains. Although research has examined self-regulated writing strategies and learners' engagement with different forms of feedback, comparatively little is known about how students describe regulating their writing when teacher and generative AI feedback are both available. In particular, limited attention has been paid to how learners compare feedback sources, resolve conflicting suggestions, and incorporate selected comments into planning, monitoring, checking, and revision. Evidence concerning Chinese non-English-major undergraduates remains especially limited.

Against this background, a process-oriented qualitative investigation can provide insight into how learners interpret and use teacher and AI-generated feedback and how they describe its relationship with their metacognitive regulation in L2 writing.

Accordingly, this study adopts a process-oriented qualitative perspective to examine how Chinese undergraduate L2 writers describe planning, monitoring, checking, and revising when using teacher and AI-generated feedback. It also investigates how participants distinguish between the two feedback sources, how they respond when the sources differ, and what learner-related or contextual factors appear to shape their reported engagement. Rather than evaluating writing scores or testing the effectiveness of a particular feedback system, the study seeks to provide an exploratory account of learners' feedback-use and metacognitive regulation experiences.

3. Methodology

3.1. Research design

This study adopted a qualitative research design to explore how Chinese undergraduate L2 writers describe their metacognitive regulation when using both teacher feedback and AI-generated feedback. A qualitative approach was selected because the study aimed to obtain detailed accounts of learners' reported planning, monitoring, checking, and revising practices rather than to test predetermined hypotheses or quantify changes in writing performance [20]. This approach is appropriate for examining context-dependent phenomena such as feedback engagement and metacognitive regulation, for which participants' interpretations and accounts of their experiences are central [21].

For the purposes of this study, metacognitive regulation was examined through four reported activities: planning, monitoring, checking, and revising. Planning referred to setting writing goals and organising ideas before or during drafting. Monitoring referred to tracking the development of the text and its alignment with task requirements. Checking referred to reviewing language, organisation, length, and task fulfilment. Revising referred to making changes on the basis of self-evaluation, teacher feedback, or AI-generated feedback.

In this study, human-AI collaborative feedback refers to students' use, comparison, and integration of teacher feedback and AI-generated feedback during drafting or revision. It does not refer to a formally integrated system in which teachers and AI jointly produce a single feedback response.

In line with this purpose, the main data collection tool was a set of semi-structured interviews. This format enabled the interviews to address predefined topics related to metacognitive regulation and feedback use while allowing participants to explain their experiences in their own words and provide specific examples [22]. In the interviews, students were asked to describe how they used AI-based writing tools, especially Doubao and ChatGPT, in drafting and revising their English texts and how the

feedback from these tools interacted with, or even contradicted, the feedback they received from their teachers.

The study did not require all participants to use the same AI tool, prompt, writing task, or sequence of feedback. Instead, it examined naturally occurring experiences of using teacher and AI feedback. This variation was treated as part of the research context and is considered when interpreting the findings.

3.2. Sampling and participants

Purposive sampling was first used to recruit students who had recent experience of working with human-AI collaborative feedback, so that they could offer detailed accounts of their writing and revision processes [20]. Table 1 shows the participant characteristics, after the initial participants had been recruited, snowball sampling was used to identify additional students who met the inclusion criteria and were willing to participate.

The inclusion criteria were that participants: (1) were Chinese undergraduate non-English majors; (2) had recent experience completing English writing tasks; (3) had received teacher feedback on their writing; and (4) had used a generative AI tool, such as ChatGPT or Doubao, to obtain feedback or revision suggestions. Students who had no recent experience of using both feedback sources were not included.

Five Chinese undergraduate non-English majors participated in the study. They came from five different fields, including education, media, Chinese language and literature, computer science and international economics and trade. All participants were taking or had recently completed compulsory College English courses involving regular English writing tasks. They reported receiving teacher feedback and using AI tools to check or revise at least some of their drafts. They had studied English for about 11-12 years. Their typical writing tasks included argumentative essays, business English reports, abstracts, and short course-related academic texts. Participants also varied in their frequency and purposes of AI-tool use. Some reported occasional use, whereas others described frequent or routine use of AI for grammar and style checking.

The participants were recruited from several universities located in different cities in China. The first group of participants was contacted through the researcher's personal academic network, including former classmates, university contacts, and student groups. After the initial participants had completed the study, snowball sampling was used, whereby they were asked to share the recruitment information with other undergraduate students who met the selection criteria.

The study was designed as a small-scale exploratory qualitative investigation. The purpose of the sample was to obtain detailed, information-rich accounts rather than to achieve statistical representativeness. However, the small number of participants limits the diversity and transferability of the findings.

Table 1. Participant Characteristics

Participants	Year level	Major	English proficiency	Gender
Julie	4	Education	Passed CET- 4	Female
Lily	4	Media	Passed CET- 4 and CET-6	Female
Anna	2	Chinese Language and Literature	Not yet taken CET- 4	Female
Jason	3	Computer Science	Passed CET-4	Male
Amy	2	International Economics and Trade	Passed CET-4	Female

Note: All participant names are pseudonyms. CET-4 and CET-6 refer to China's College English Test Bands 4 and 6.

3.3. Data collection

Data were collected through individual semi-structured interviews. Semi-structured interviews were chosen because they allowed the researcher to explore pre-identified themes related to metacognitive regulation and feedback use, while also giving participants freedom to elaborate on their own experiences and examples [23].

All interviews were conducted online using Tencent Meeting, as the participants were based in different locations. Before each interview, participants received an information sheet explaining the purpose and procedures of the study and provided informed consent. At the beginning of each session, the researcher restated the purpose of the study, confirmed participants' voluntary participation, and

explained that pseudonyms would be used and identifying information would be removed from the reported data. The interviews were carried out in Chinese so that students could describe their experiences in as much detail and nuance as possible. Under the relevant institutional policy, formal ethics committee approval was not required for this minimal-risk student interview study. Nevertheless, informed consent, voluntary participation, confidentiality, and the right to withdraw were explained to all participants.

An interview guide was used to ensure that the key topics were covered. These included participants' recent L2 writing tasks; their experiences of receiving and using teacher feedback and AI-generated feedback; their reported planning, monitoring, checking, and revising practices; and their perceptions of the benefits and challenges of using both feedback sources. The interview questions were informed by the study's conceptualisation of metacognitive regulation. Questions concerning goals, outlines, and idea organisation addressed planning; questions concerning rereading, problem awareness, and alignment with task requirements addressed monitoring and checking; and questions concerning the acceptance, rejection, adaptation, or extension of feedback addressed revision. Participants were also asked how they responded when teacher and AI feedback differed. Follow-up questions and prompts were used when clarification or specific examples were required. Each interview lasted approximately 30-45 minutes. All interviews were audio-recorded with the participants' permission and transcribed verbatim for analysis.

3.4. Data analysis

Data were analysed using reflexive thematic analysis following Braun and Clarke (2019) [24]. This approach was appropriate because it allowed the researcher to identify patterns of shared meaning across participants' accounts while recognising that themes were developed through an interpretive engagement between the researcher and the data.

The analysis went through several overlapping phases. First, the researcher became immersed in the data by reading and re-reading the transcripts of the interviews and any accompanying field notes. Second, initial codes were generated manually by identifying segments related to planning, monitoring, checking, revision, and engagement with teacher and AI feedback. Both deductive codes, informed by metacognitive regulation and feedback literature, and inductive codes, emerging from participants' own language, were used. Deductive coding used planning, monitoring, checking, and revising as sensitising concepts rather than as fixed categories into which all data had to be placed. Inductive coding was used to capture participants' own descriptions, including trust in teacher judgement, AI use for rapid language checking, selective acceptance of suggestions, and uncertainty when the two feedback sources conflicted.

Third, coded extracts were collated and potential themes were developed by grouping related codes, for example codes concerning "using AI for quick checking" or "trusting teacher judgement more than AI" contributed to candidate themes concerning the perceived functions and authority of different feedback sources. Fourth, these themes were reviewed in relation to the coded data and the full data set, and refined to ensure internal coherence and clear distinction between themes. Fifth, themes were defined and named, with attention to how they captured different aspects of learners' metacognitive regulation under human-AI collaborative feedback. Finally, illustrative quotations were selected and an analytic narrative was written to show how the themes addressed the research questions.

To enhance analytical transparency, the researcher also considered whether the themes adequately represented both shared patterns and differences across participants' accounts.

3.5. Researcher Reflexivity

The researcher approached the study with prior knowledge of L2 writing instruction and an interest in AI-assisted feedback. These experiences may have influenced the interview prompts, coding decisions, and interpretation of participants' accounts. To remain aware of this influence, the researcher reflected on personal assumptions and documented key analytical decisions during the coding and theme-development process. The themes should therefore be understood as interpretive constructions developed through engagement among the researcher, the participants' accounts, and the conceptual framework.

4. Findings

Drawing on interview data, the findings are organised around the three research questions. They show (1) how Chinese undergraduate L2 students described their engagement in planning, monitoring, checking and revising when they write; (2) how participants perceived the relationship between metacognitive regulation and their writing processes and performance; and (3) how students perceive

and differentiate between teacher feedback and AI feedback within a human-AI collaborative feedback ecology.

4.1. Students' engagement in planning, monitoring, checking and revising

All five participants said that they did some kind of planning, monitoring, checking and revising, but how often and how deeply they did these things varied considerably. Three of them (Julie, Anna and Amy) usually planned their texts before they started to write, whereas Lily and Jason preferred to begin writing and adjust their ideas as they went. All participants described at least some level of self-checking during composing. However, only Lily, Jason and Amy reported more systematic, paragraph-by-paragraph monitoring and going through several rounds of revision. Julie and Anna, in contrast, tended to make relatively limited changes that were very close to what teachers or AI had already pointed out.

When it came to planning, Julie and Anna both used outlines primarily as a means of keeping the text organised and not going off topic. "If I don't do a simple outline first, I easily go off topic later" said Julie, and Anna explained that an outline "helps me fix the main point and the role of each paragraph, otherwise the length and content can get messy". Similarly, Amy adopted the most deliberate planning routines, describing how a detailed outline with the length and purpose of each paragraph helped her, in her view helped her to keep the argument tight and control the length. By contrast, Jason preferred to write with only a few keywords, adjusting ideas as he went along, and Lily planned largely in her head.

Monitoring and checking during writing showed similarly diverse patterns. Lily reported very frequent checking, stating that:

Sometimes I have not even finished a sentence, and I already read it again to check the tense and word form. I then reread longer sections to consider sentence order and coherence.

Jason also noted that his self-checking frequency had increased: "When I finish a paragraph, I check it myself once, then let AI check it again." Amy described pausing several times to reread and deliberately checking logic, language and length against both the outline and the task prompt. Julie reported moving from checking only at the end of the draft to rereading each paragraph to ensure that she was still addressing the task. By contrast, Anna stated:

The way I check has remained largely the same. I mainly check whether I have answered the question and whether the length is appropriate. I seldom check grammar while writing.

Overall, all participants reported engaging in planning, monitoring, checking and revising to some extent, but only a subset described these activities as systematic and strategically coordinated, particularly when comparing and combining teacher and AI feedback.

4.2. Participants' perceptions of the relationship between metacognitive regulation and L2 writing

Most participants reported that their awareness of planning, monitoring and revising had increased through repeated engagement with feedback and that they now paid more deliberate attention to how they organised and checked their writing. Three students (Julie, Jason and Amy) explicitly connected these changes in planning, monitoring and revision to perceived improvements in their writing, especially in terms of staying on topic, organising ideas more clearly and reducing language errors. By contrast, Lily and Anna felt that any perceived influence was limited or confined to specific aspects.

For some, feedback experiences were associated with clearer and more explicit writing goals. Jason reported that AI comments on repetition and simple sentence patterns had changed what he attended to during writing:

Before, my goal was just to finish the task. Now I consciously think about using more varied sentence patterns and fewer very simple sentences.

Amy said that her repeated difficulties with long texts and wandering ideas made her more aware of the need for self-regulation in her writing. These experiences led her to set clearer objectives for herself-to keep the length in check, to narrow the main points, and to check that the content, language and format were all up to the expected standard. Julie reported a similar change. She felt that teacher comments on topic focus, as well as AI repeatedly flagging spelling and form issues, had increased her awareness of details and encouraged her to check each paragraph more carefully against the task requirements.

On this basis, participants perceived that changes in monitoring and checking supported their writing. Lily explained that deliberately switching between a "teacher view" and an "AI view" when rereading helped her identify both structural problems and language issues more clearly:

Sometimes I pretend I am the teacher and look at the structure and logic; sometimes I pretend I am

the AI and only look for grammar and collocation problems.

Amy described a similar strategy, stating that when she reread her writing, she alternated between a "teacher view" and an "AI view," focusing on logic and genre from one perspective and language accuracy from the other. Julie also felt that moving from end-of-draft checking to paragraph-level checking had helped her reduce off-topic writing.

Three participants (Lily, Jason and Amy) reported revising beyond the specific points identified in feedback. Lily explained:

If I feel the teacher's comment makes sense, I will follow the idea and make additional changes, rather than only fixing that one sentence.

Jason commented that, although AI had identified only a few repeated verbs, he decided to revise similar repetition throughout the text because he wanted to improve the overall style rather than simply correct the explicitly marked problems.

Conversely, Anna felt that metacognitive regulation had influenced her writing only slightly. She continued to depend mainly on teacher guidance, rarely monitored her language while drafting, and tended to revise only what had been explicitly marked. As she explained:

I usually just correct the places the teacher circles, and when I am busy, I sometimes correct only two spelling mistakes and leave the other suggestions unchanged. The way I check has not changed much; I pay more attention to clear grammar errors because they lose the most points.

Her account indicates a limited reported shift in how she planned, monitored and revised her writing. It also suggests that access to both teacher and AI feedback did not automatically lead to deeper engagement, particularly when assessment concerns and time constraints remained dominant.

Overall, most participants perceived that more deliberate planning, monitoring, checking and revision supported topic focus, organisation and language accuracy. However, these perceived benefits were uneven. Participants who described more active comparison and extension of feedback also reported broader changes in their writing practices, whereas those who focused mainly on explicitly marked, score-related errors described more limited change. The interview data do not establish whether these differences were related to participants' writing proficiency, because writing ability was not independently assessed.

4.3. Students' perceptions of teacher and AI feedback

All five participants clearly differentiated between AI and teacher feedback, although their views of the roles and authority of the two sources varied. Four of them (Julie, Lily, Anna and Amy) regarded teacher feedback as the primary and more reliable source, especially for content, argumentation, structure, assessment criteria and genre conventions. In contrast, AI feedback was generally regarded as useful for checking grammar, spelling, collocations and sentence-level phrasing. Jason expressed a more balanced view and sometimes preferred AI-generated stylistic suggestions to the teacher's preference for simpler wording.

Most participants regarded teachers as the primary authority for higher-order and assessment-related feedback. Students said teachers had a better sense of the assignment, and could give more targeted and context-sensitive feedback. They also said that teachers were able to identify errors and diagnose more serious problems with ideas and organisation that they themselves might easily miss.

Julie, for instance, recalled that in an essay on urban environment:

The teacher told me that the focus should be on air quality and problems such as traffic and housing, rather than on general suburban issues. This made the direction of the essay much clearer in later drafts.

Anna similarly commented that teachers pointed out "off-topic main ideas and unclear logic," whereas "if I just give my essay to AI, as long as there are no major grammar mistakes, it thinks the essay is fine." Amy emphasised that in business English writing, "only the teacher can really judge whether my style is formal enough and appropriate for the professional context."

By contrast, all participants associated AI feedback mainly with language form. For example, Julie reported "using AI every time after I finish a draft to fix obvious things like tense, singular-plural and spelling". Jason stated that he basically cannot write English without AI now, because it points out repeated sentence patterns and wrong collocations, and gives alternative sentences. Amy explained:

AI as useful for basic language checking, such as scanning grammar, flagging awkward collocations and suggesting improved linking sentences, particularly for formal assignments.

Although all participants acknowledged the practical value of both sources, four expressed a clear preference for teacher feedback when the two sources differed. They explained that teachers were familiar with course expectations and grading criteria and could provide more context-sensitive feedback on content, logic and genre.

As Anna said:

If AI and the teacher say different things, I always listen to the teacher because the teacher is familiar with the exam content and the grading criteria.

Some participants also noted that teachers ultimately determined their grades, which increased the weight assigned to teacher feedback in their revision decisions.

Participants' perceptions of AI were mixed rather than uniformly negative. Most valued its efficiency and breadth, particularly its ability to identify easily overlooked language problems and suggest alternative expressions, while remaining cautious about its perceived limitations.

Julie worried that advanced words suggested by AI could make her writing sound inauthentic. She explained:

The teacher would immediately see they are from AI, and I cannot really remember or control those words", so she "usually does not use them.

Lily described AI as "useful, but still a black box" and stated that "you cannot fully trust it, especially for important things like references." Anna was the most sceptical participant, stating:

I did not really like this combination because many AI suggestions did not match what we were taught in class. For me, the main benefit was catching a few basic grammar errors.

Overall, most participants viewed teacher and AI feedback as complementary rather than interchangeable. Teacher feedback was generally assigned greater authority in relation to content, logic, genre and assessment expectations, whereas AI feedback was used primarily for language-level checking and alternative expression. Some participants also used AI-generated suggestions for stylistic revision, although writing confidence was not independently assessed in this study.

5. Discussion

This study investigated how Chinese undergraduate students described regulating their L2 writing when using teacher and AI-generated feedback, and how they perceived the roles of these two feedback sources. The findings suggest that students assigned distinct roles to the feedback from teachers and artificial intelligence, and there were significant differences in their participation levels during the planning, monitoring, and revision processes. Moreover, only some participants reported extending feedback beyond explicitly marked problems and incorporating it into broader changes in their writing practices.

A key finding is that the participants did not generally regard teacher feedback and AI-generated feedback as equally authoritative sources of guidance. They described a clear hierarchy of functions, believing that teacher feedback is more trustworthy and authoritative in terms of content, argumentation, structure, and genre. While AI feedback was used mainly as an auxiliary tool for checking and rewriting articles at the language level. This finding is consistent with Hyland and Hyland (2006), who emphasized the authority of teacher comments and its close connection with the assessment expectations in L2 writing contexts [9].

The present study extends this discussion by illustrating how perceived teacher authority operated when participants compared and selected feedback from different sources during drafting and revision. One possible reason is that teachers are associated with assessment criteria and final scores, and are considered to have a better understanding of task requirements; while AI is related to speed and coverage, but also has issues such as limited contextual understanding and occasional errors.

This functional hierarchy also suggests that the perceived authority of feedback was shaped not only by its content but also by its institutional source. Teacher comments carried additional weight because teachers interpreted task requirements and controlled assessment outcomes, whereas AI feedback had no formal role in grading. This may help explain why several participants followed teacher judgement when the two sources conflicted, even when they recognised the convenience or stylistic usefulness of AI-generated suggestions.

Another important finding is that participants' metacognitive regulation appeared to shape how they interpreted and acted upon teacher and AI-generated feedback. Not all students who use both types of feedback sources report significant progress. Students who set clear goals and monitor their texts more closely tend to make stronger use of feedback. They also perceived greater benefits in relation to topic focus, organisational clarity, and language accuracy.

This interpretation is consistent with the view that planning, monitoring, and revision are central components of self-regulated L2 writing [25]. However, the present findings add that access to multiple feedback sources did not automatically produce more strategic regulation. Some participants used the coexistence of teacher and AI feedback as a resource for monitoring their own writing. For example, they alternated between a "teacher perspective" and an "AI perspective" during re-reading or extended feedback from marked locations to similar unmarked problems elsewhere in the text. Other participants remained more strongly oriented towards scores and explicit correction and reported little change in

how they planned or monitored their writing.

The contrast between participants who extended feedback and those who corrected only explicitly marked errors is also relevant to feedback literacy. Participants who compared sources, judged the relevance of suggestions, and applied a comment to similar problems elsewhere demonstrated more active engagement with feedback. By contrast, participants who relied mainly on teacher authority or corrected only score-related errors appeared to treat feedback more narrowly as information to be followed. This difference suggests that the educational value of human-AI collaborative feedback may depend partly on learners' ability and willingness to evaluate and transform feedback into self-directed revision.

At the same time, the differences among participants should not be attributed solely to metacognitive regulation. Their reported practices may also have been shaped by assessment pressure, time constraints, prior writing experience, English proficiency, familiarity with AI tools, and the types of writing tasks they completed. For example, Anna's emphasis on score-related grammar errors suggests that assessment priorities may restrict broader engagement, while Jason's frequent AI use may have increased his willingness to consider AI-generated stylistic alternatives. These interpretations remain tentative because the study did not independently measure proficiency, feedback literacy, AI literacy, or writing confidence.

Taken together, the findings suggest that, within this small sample, teacher feedback anchored most participants' higher-order and assessment-related writing decisions, whereas AI-generated feedback was used primarily for language-level support and, in some cases, stylistic revision. Human-AI collaborative feedback may provide opportunities for metacognitive regulation, but participants' engagement with these opportunities varied according to their reported goals, feedback-use practices, and perceptions of the two sources.

Realising this potential may therefore require explicit support for feedback literacy. Learners may need opportunities to compare comments from different sources, evaluate their relevance to the task, justify which suggestions they accept or reject, and connect feedback with their own planning, monitoring, checking, and revision.

The contribution of this study is therefore not to demonstrate that human-AI collaborative feedback improves writing performance, but to provide an exploratory account of how a small group of Chinese undergraduate L2 writers interpreted, prioritised, and acted upon two feedback sources. The findings highlight the distinction between access to feedback and strategic engagement with feedback: receiving both teacher and AI comments did not by itself ensure broader revision or stronger metacognitive regulation. This process-oriented perspective complements outcome-focused research by showing how learners' reported decisions may shape the use of available feedback.

6. Conclusion

This study explored how Chinese undergraduates described regulating their L2 writing under human-AI collaborative feedback, how this regulation was perceived to relate to their writing processes and performance, and how they view different feedback sources.

Regarding RQ1, the results indicated considerable variation in the ways students planned, monitored, checked and revised their work. Some students described relatively systematic practices involving outlining, paragraph-level monitoring, multiple rounds of revision, and the comparison of teacher and AI feedback. Others planned very little and limited their revisions to points made explicitly salient by the teacher or the AI system, with only weak evidence of systematic self-checking.

For RQ2, the results suggest that participants perceived metacognitive regulation as supporting writing through clearer goals, more focused monitoring, and broader revision. Students who reported clearer goals, closer monitoring, and revision beyond explicitly marked problems also described better perceived control over topic focus, organisation, and language accuracy. In contrast, those who are more concerned about not losing scores and trying to correct superficial errors only reported moderate changes.

In this sense, human-AI collaborative feedback may provide opportunities for metacognitive regulation, but the extent to which participants engaged with these opportunities varied according to their goals, feedback-use practices, and willingness to evaluate and extend feedback.

For RQ3, the results show that students perceived teacher and AI-generated feedback as serving different functions and carrying different levels of authority. Teachers are considered the main authority on content, logic, evaluation criteria and genres, while artificial intelligence is valued for its speed and help with language level problems. If the two sources are inconsistent, most participants rely on the teacher's judgment, mainly using artificial intelligence to identify small errors and propose alternative words. At the same time, some learners show behaviours consistent with developing feedback literacy, including selectively accepting, adapting, or rejecting AI-generated suggestions rather than following

them automatically.

The findings have implications for instruction. Writing teachers may need to be more explicit about the respective purposes and limitations of teacher and AI-generated feedback, for example by clarifying which aspects of writing each source is intended to support. They can also design tasks that ask students to compare comments from the two sources, justify the suggestions they decide to act on, and record how feedback actually shapes their planning and revising. Furthermore, explicit classroom attention to feedback literacy and metacognitive regulation may encourage students to move beyond isolated error correction towards more strategic and reflective feedback use. Although these implications are potentially useful, they should be interpreted cautiously because they are derived from a small exploratory sample and were not evaluated through classroom intervention or writing-performance measures.

This research also has some limitations. First, the study was based on a small qualitative sample of five participants from [one university/several universities], which limits the diversity and transferability of the findings. Second, the data consisted mainly of self-reported interview accounts and were not triangulated with students' drafts, teacher comments, AI prompts and outputs, revision histories, or classroom observations. The findings therefore reflect participants' perceptions rather than directly observed regulatory behaviour or objectively measured writing improvement.

Third, participants differed in their use of AI tools, prompts, writing tasks, and teacher feedback, meaning that human-AI collaborative feedback was not a fully standardised condition. Fourth, the interviews were conducted in Chinese and reported in English, which may have resulted in some loss of nuance during translation. Finally, the researcher's role in interviewing, coding, and interpreting the data may have influenced the themes developed.

Future research could involve larger and more diverse participant groups and adopt mixed-method or longitudinal designs combining interviews with draft histories, teacher comments, AI interaction records, classroom observations, or writing-process data. Such research could examine how different feedback configurations relate to metacognitive regulation and writing development over time and could distinguish more clearly between participants' perceived improvement and changes demonstrated in their actual writing.

Overall, this study provides an exploratory qualitative account of how five Chinese undergraduate L2 writers interpreted, prioritised, and used teacher and AI-generated feedback. The findings suggest that access to both feedback sources did not automatically produce deeper metacognitive regulation. Rather, participants differed in the extent to which they compared feedback, evaluated its relevance, and extended it into self-directed revision. The study therefore highlights the importance of distinguishing between the availability of feedback and learners' strategic engagement with it.

References

1. Ministry of Education of the People's Republic of China, 2020. College English Teaching Guidelines. Beijing: Higher Education Department, Ministry of Education. Available at: <http://www.moe.gov.cn/>
2. Hyland, K., 2003. Second Language Writing. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511667251>
3. Hyland, K., 2019. Second Language Writing. 2nd ed. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108693974>
4. Zimmerman, B.J., 2002. Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), pp. 64-70. https://doi.org/10.1207/S15430421TIP4102_2
5. Flower, L. and Hayes, J.R., 1981. A cognitive process theory of writing. *College Composition and Communication*, 32(4), pp. 365-387. <https://doi.org/10.2307/356600>
6. Chen, C.F.E. and Cheng, W.Y.E.C., 2008. Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes.
7. Hattie, J. and Timperley, H., 2007. The power of feedback. *Review of Educational Research*, 77(1), pp. 81-112. <https://doi.org/10.3102/003465430298487>
8. Bitchener, J., Young, S. and Cameron, D., 2005. The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), pp. 191-205. <https://doi.org/10.1016/j.jslw.2005.08.001>

-
9. Hyland, K. and Hyland, F., 2006. Feedback on second language students' writing. *Language Teaching*, 39(2), pp. 83-101. <https://doi.org/10.1017/S0261444806003399>
 10. Lee, I., 2017. *Classroom Writing Assessment and Feedback in L2 School Contexts*. Singapore: Springer. <https://doi.org/10.1007/978-981-10-3924-9>
 11. Attali, Y. and Burstein, J., 2006. Automated essay scoring with e-rater® V.2.0. *Journal of Technology, Learning, and Assessment*, 4(3), pp. 1-29.
 12. Li, J., Link, S. and Hegelheimer, V., 2015. Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, pp. 1-18. <https://doi.org/10.1016/j.jslw.2014.10.004>
 13. Dikli, S., 2006. An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment*, 5(1), pp. 1-36.
 14. Ferris, D.R., 2003. *Response to Student Writing: Implications for Second Language Students*. Mahwah, NJ: Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410607201>
 15. Zhang, Z., Aubrey, S., Huang, X. and Chiu, T.K., 2025. The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*, pp.1-19.
 16. Li, S., 2025. Generative AI and second language writing. *Digital Studies in Language and Literature*, 2(1), pp.122-152.
 17. Ranalli, J., Link, S. and Chukharev-Hudilainen, E., 2017. Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, 37(1), pp. 8-25. <https://doi.org/10.1080/01443410.2015.1136407>
 18. Carless, D. and Boud, D., 2018. The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), pp. 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
 19. Zimmerman, B.J., 2000. Attaining self-regulation: A social cognitive perspective. In *Handbook of self-regulation* (pp. 13-39). Academic press.
 20. Creswell, J.W., 2014. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks, CA: SAGE.
 21. Dörnyei, Z., 2007. *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.
 22. Richards, K., 2003. *Qualitative Inquiry in TESOL*. Basingstoke: Palgrave Macmillan. <https://doi.org/10.1057/9780230505056>
 23. Adeoye-Olatunde, O.A. and Olenik, N.L., 2021. Research and scholarly methods: Semi-structured interviews. *Journal of the american college of clinical pharmacy*, 4(10), pp.1358-1367.
 24. Braun, V. and Clarke, V., 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), pp. 589-597. <https://doi.org/10.1080/2159676X.2019.1628806>
 25. Teng, L.S. and Zhang, L.J., 2016. A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal*, 100(3), pp. 674-701. <https://doi.org/10.1111/modl.12339>