

**AI-Calibrated Metacognition: How Genre-based ChatGPT Feedback and Interaction
Shape L2 Writers' Metacognitive Judgments and Self-regulation**

Issam Rian

Second Language Acquisition and Teaching (SLAT)

The University of Arizona

Abstract

Large language models (LLMs) increasingly mediate feedback in L2 writing, yet little is known about how such output reshapes learners' decision-making. This qualitative study examines how genre-based ChatGPT feedback and dialogue shape novice L2 writers' metacognitive judgments (MJs)—their basis, accuracy, and evolution—and how these judgments drive self-regulated learning (SRL). In a first-year composition course for international students, nine participants completed three genre-based assignments and engaged in structured, AI feedback cycles using a custom GPT, *Genre Guru*, designed to provide genre-specific feedback. Data sources included AI-feedback reflections, ChatGPT interaction logs, end-of-course reflections, and five post-semester interviews. Using thematic analysis, MJs were traced across the four domains of genre-specific knowledge (formal, rhetorical, process, subject-matter) and mapped onto the phases of SRL (forethought, performance, self-reflection). Four themes emerged: (1) students moved from initial skepticism to measured trust; (2) learners critically evaluated AI suggestions, preserving text ownership; (3) writers integrated the four domains and articulated genre awareness; and (4) emotions and motivation drove self-regulatory cycles. Findings suggest that, when framed by genre theory and embedded in reflective dialogue, LLM-mediated feedback can function as a catalyst for *AI-calibrated metacognition*: enhancing judgment quality, scaffolding regulation, and expediting adaptive, autonomous L2 writing without displacing human agency.

Keywords: AI; decision-making; metacognition; genre knowledge.

Introduction

Large language models (LLMs) could transform second language acquisition and teaching, yet the pace of generative AI raises questions about whether it can match human expertise. While AI can simulate certain cognitive functions, it does not currently demonstrate self-awareness or subjective experience. One differentiating capacity, central to autonomous and reflective learning, is metacognition: thinking about one's own thinking (Flavell, 1979). Humans can monitor their cognitive processes, update them, and take consequential action to reach their goals. In educational contexts, these processes are described as metacognitive monitoring and control: agentic, goal-directed regulation grounded in one's model of tasks, goals, and the world—capacities that current AI systems only indirectly approximate.

In L2 writing, students' engagement with AI-generated corrective feedback is salient. The leap to LLMs, such as ChatGPT, marks an evolution beyond traditional Automated Writing Evaluation (AWE) systems. By generating either holistic scores, detailed comments, or both, AWEs offer learners a chance to revise their work multiple times at a pace faster than traditional human feedback (Warschauer & Grimes, 2008; Li, Link, & Hegelheimer, 2015). However, LLMs not only provide immediate, context-specific feedback on writing tasks; they also allow for interactive, real-time conversations that traditional AWEs cannot sustain. Crucially, LLMs can be prompted to provide genre-based feedback, including rhetorical considerations. This capability enables instructors and researchers to operationalize theory, such as Tardy's (2009) genre knowledge model, to provide learners with holistic, theory-grounded corrective feedback. Yet research on the integration of generative AI (GenAI) as a feedback tool in L2 writing instruction remains nascent. Little is known about how GenAI shapes novice L2 writers' metacognitive apparatus, decision-making, and chains of reasoning as they interact with genre-

based AI feedback. This study focuses on L2 writers' metacognitive judgments (MJs; Koriat, 1997)—assessments of one's own cognitive processes—and how these judgments shape feedback and self-regulatory strategies, with implications for the development of genre expertise.

Conceptual and Theoretical Frameworks

Metacognitive Judgments (MJs)

Researchers describe the subjective experience of metacognition as consisting of three components: metacognitive knowledge, metacognitive monitoring, and metacognitive control (Dunlosky & Metcalfe, 2008). On the one hand, metacognitive knowledge is domain-independent and refers to knowledge that people can consciously articulate about their own mental abilities. This type of knowledge can be declarative (a writer's knowledge of their skills), procedural (a writer's knowledge of how to use their skills), or conditional (a writer's knowledge of when and why to apply those skills) (Schraw & Dennison, 1994). On the other hand, metacognitive monitoring and control—collectively called metacognitive *regulation*—are domain-specific. For a particular cognitive skill, people can evaluate their progress (monitoring) or devise and implement strategies to reach their goals for that skill (control).

When judgments are made at the meta-level about one's own cognitive processes, they are metacognitive judgments (MJs)—self-assessments that students make about their learning (in this study: writing development and genre knowledge). Three dimensions of MJs are the focus of this study: basis, accuracy (Dunlosky & Thiede, 2013), and evolution. Although the basis of students' judgments is rooted in factors that shape their writing skills, the quality of those judgments may not correlate with actual performance; consequently, accuracy can vary. Since most research on MJs uses experimental designs that rely on immediate retrieval, this study adds

a temporal dimension: how L2 writers' MJs change over time across three genres. This dimension is important in AI-mediated learning because it underscores the dynamic, subjective nature of metacognitive awareness and its influence on writing behavior.

Self-regulation

Once writers have made their MJs, they need to regulate their behavior, calling on strategies and techniques to improve their writing outcomes, thereby engaging in self-regulated learning. This study adopts Zimmerman and Moylan's (2009) model of self-regulation, which comprises three phases—forethought, performance, and self-reflection (Figure 1)—to capture the strategies L2 writers deploy following their MJs. Self-regulation unfolds as a cycle: in the forethought phase, learners analyze the task through goal setting and strategic planning, grounded in motivational and strategic foundations. During performance, they oscillate between self-control and self-observation, enabling real-time adjustments that keep their work aligned with the plan. In self-reflection, they evaluate results against standards and attribute causes, shaping self-reactions and adaptive or defensive inferences. These beliefs and emotions feed into the next forethought, completing the loop.

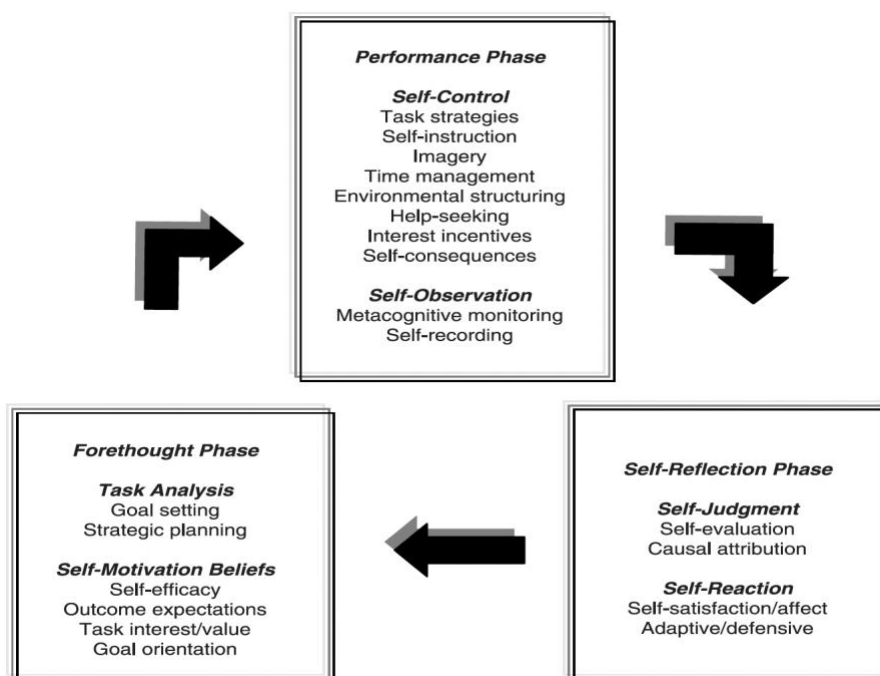


Fig. 1. Zimmerman and Moylan's (2009) Self-regulation Model

Tardy et al.'s (2020) Model for Genre Knowledge

Tardy's (2009) framework was the first to conceptualize genre knowledge as integrated within writing rather than as an isolated skill. Early theories of writing treated genre knowledge as an independent domain to be mastered in isolation. Hayes's (1996) model, for example, distinguishes genre knowledge from knowledge of topic, audience, and language—all of which reside in long-term memory. However, Tardy (2009) argued that, to write proficiently in a genre, writers must draw on four domains of knowledge: (1) formal knowledge, pertaining to textual form, lexicogrammatical features, and content; (2) rhetorical knowledge, including purpose and genre-specific rhetorical conventions; (3) process knowledge: how the genre is produced; and (4) subject-matter knowledge, i.e., discipline-specific knowledge (p. 21).

Tardy et al. (2020) introduced a framework that expands Tardy (2009) to clarify relationships among genre knowledge, genre awareness, metacognition, and conditional

knowledge (Figure 2). The framework makes a major change: renaming the construct genre knowledge as genre-specific knowledge. In this model, genre-specific knowledge and genre awareness constitute a writer's overall genre knowledge. A second feature is its assumption of multilingual writers as the norm, which the authors deem imperative for capturing linguistic diversity in a world where most people are at least bilingual.

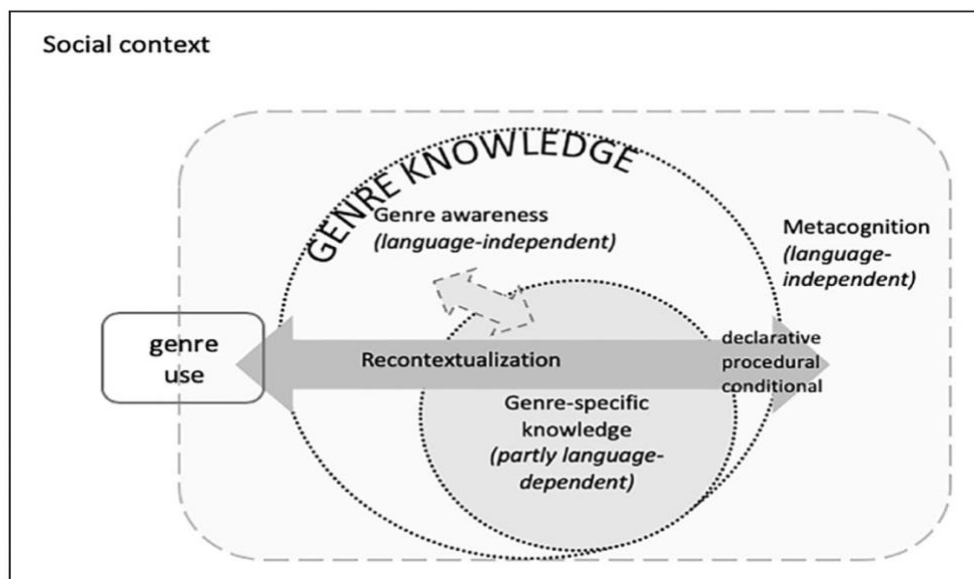


Fig. 2. Tardy et al.'s (2020) new framework for genre knowledge

Literature Review

AI Feedback, Metacognition, and Self-regulation in L2 writing

Recent work has begun to examine how L2 writers' metacognitive processes interact with AI-generated feedback. Several studies highlight metacognition's role in facilitating L2 writing with LLMs. Yao et al. (2024) conducted a qualitative study tracking 13 postgraduate students as they composed sections of research papers and observed frequent use of planning, monitoring, evaluating, and information-management strategies; *debugging*—how writers respond to ChatGPT's inaccuracies—emerged as a context-specific metacognitive dimension, including

refining prompts, revising outputs, and consulting external sources (e.g., Google) to verify accuracy.

Affect and metacognition appear closely intertwined in AI-mediated L2 writing. In a 15-week course with 19 Korean undergraduates, Lee (2024) found that groups restricted to consulting ChatGPT only for Draft 3 (after peer and instructor feedback on Drafts 1–2) collaboratively monitored and controlled their process and showed increased confidence and persistence. Shen and Tao (2025) reported that Chinese EFL writers with higher metacognitive awareness and AI-based self-efficacy experienced less anxiety; AI-based self-efficacy fully mediated the planning–anxiety relationship. Focusing on direct AI feedback, Mohammed and Khalid (2025) randomly assigned 322 EFL learners to AI-feedback vs. delayed teacher-feedback conditions over ten weeks of short expository essays; AI feedback improved proficiency and enhanced emotional regulation, motivation, and Foreign-Language Peace of Mind (FLPoM), boosting trait emotional intelligence while reducing anxiety. Students valued the nonjudgmental, neutral tone, fostering calm and encouraging self-reflection and ownership.

Research on self-regulation likewise reports benefits. In a study using Zimmerman’s (2000) model, with surveys ($n = 55$) and think-alouds ($n = 8$), Kamelia and Linda (2023) concluded that AI use enhanced efficiency across phases (improved planning, increased follow-up questioning, and sharper evaluation of lower-level issues), fostering autonomy and reflection. Aladini et al. (2025) similarly found that AI feedback increased proficiency and promoted mindfulness as measured by the Mindful Attention Awareness Scale (MAAS), facilitating self-regulatory actions. Fathi and Rahimi (2024) varied the explicitness of ChatGPT’s feedback over eight sessions with 14 Iranian CEFR B1 learners and, drawing on Vygotskian theory, showed

that learners progressively relinquished the need for explicit guidance, moving through the zone of proximal development (ZPD) toward independent control.

In parallel experimental designs, Apriani et al. (2024) and Teng (2024) found that immediate, personalized AI feedback heightened error awareness and bolstered confidence and perceived competence. At scale, Jin et al. (2025) surveyed 1,073 UK-based EFL postgraduate students and, using structural equation modeling (SEM), showed that self-efficacy predicted self-regulation, with AI self-efficacy exerting a stronger effect; use of AI for “transformation”—deeper idea generation, critical thinking, and reflection—was associated with greater critical engagement, higher writing quality, and higher perceived utility value of writing.

AI Feedback, Decision-making, Self-assessment in L2 Writing

Student engagement remains underexplored vis-à-vis generative AI. Ranalli (2021), in the context of AWEs, showed that writers who understand Grammarly’s affordances and limits develop “calibrated trust.” Extending this line to LLMs, three studies—Yan & Zhang (2024), Koltovskaia et al. (2024), and Ma et al. (2025)—used Ellis’s (2010) cognitive-behavioral-affective framework. Cognitively, L2 writers noticed, reread, compared, set goals, and dismissed misaligned feedback; such gains may depend on proficiency and technological savviness (the lower-proficiency pair in Yan & Zhang, 2024, struggled with self-regulation). Behaviorally, students preferred local over global feedback and sometimes copied paraphrases or mechanical fixes (Koltovskaia, 2024). Unlike Yan & Zhang (2024) and Koltovskaia (2024), Ma et al. (2025) observed students actively seeking supplementary input (peers, instructor, online sources). Across studies, iterating and refining prompts was indispensable. Affective responses were broadly positive (accuracy, convenience, lower-level fixes), but turned negative when revision felt time-consuming (Yan & Zhang, 2024), feedback was inaccurate/over-general (Koltovskaia,

2024), or prompts were misunderstood and content impractical (Ma et al., 2025); more interaction and better prompts often reversed this.

A small cluster zooms in on cognitive and behavioral uptake. Chen et al. (2024) trained 45 Chinese undergraduates to elicit ChatGPT feedback on a 250-word argumentative essay and had them journal why they accepted, contested, or ignored comments; students adopted form-focused feedback more often yet more frequently challenged content feedback. Reframing with UTAUT, Chen et al. (2025) found rejection stemmed from overload, misalignment with external sources, misinterpretation of author intent, or generic/irrelevant advice; rejection reflected perceived usefulness and a preference for teacher/peer feedback more than negative emotion.

Within the Chinese context, Yan (2024) examined feedback-seeking behavior across three cases. High-proficiency writers (Toby, Isabella) crafted detailed prompts; Lisa (moderate language, high tech competence) relied on brief prompts/exchanges, limiting reflection. Crucially, interaction with ChatGPT required metacognitive regulation as students articulated affective and cognitive judgments (p. 13).

Direct self-assessment complements this picture. Building on Nicol's (2021) *internal feedback*, Tam (2025) followed 48 Chinese EFL undergraduates over 14 weeks (diaries, two interviews, weekly chat-log exports). Students generated internal feedback by judging ChatGPT's suggestions, reflecting on improvement areas, and self-assessing through the AI; response quality hinged on personal goals, self-regulation effectiveness, and L1 use. Monitoring and control elicited better AI responses—and, reciprocally, interaction with ChatGPT generated internal feedback that drove self-regulation and autonomy (Tam, 2025, p. 231).

Metacognition and Genre Knowledge in L2 Writing

Metacognition and genre knowledge are mutually reinforcing. Scaffolding metacognition cultivates genre knowledge: L2 doctoral students who reflect on their writing milieu and genre knowledge report a readiness to employ genres as a writing tool (Negretti & McGrath, 2018). Metacognitive knowledge—especially conditional knowledge—facilitates task comprehension and develops rhetorical awareness; with journaling, attention to audience and purpose increases over time, enabling better self-assessment of responses to rhetorical demands (Negretti, 2012). Explicit teaching of metacognition through a genre-analysis approach can also foster domain integration: in an ESP course for 24 STEM doctoral students, post-course interviews converged on four intertwined dimensions— (1) genre analysis, (2) reader awareness, (3) rhetorical strategies, and (4) writing process—reflecting development toward long-term expertise (Negretti, 2021). Conversely, teaching genre analysis develops metacognitive knowledge: eight L2 pre-service teachers improved their declarative and procedural knowledge of academic genres and adopted new metaconcepts (audience, purpose, structure), though conditional knowledge, the facet most predictive of improvement, did not emerge for all participants (Negretti & Kuteeva, 2011).

Metacognition may also distinguish experts from neophytes. In a large, multi-institutional study, metacognitive reflection promoted nuanced genre awareness—an understanding that goes beyond form to rhetorical aspects—which correlated with writing for authentic audiences, a key pathway for integrating Tardy’s (2009) four genre-knowledge domains and a potential hallmark of mastery (Driscoll et al., 2019). Developing genre (including rhetorical) knowledge appears tied to self-regulation: as graduate L2 writers evaluate their work, they better monitor and control alignment with task demands, creating a positive learning loop tied to both metacognitive depth and metacognitive alignment—the extent to which writers believe their texts account for genres’

social and contextual dimensions (Negretti, 2017). Expert STEM scientists likewise exhibit sophisticated self-regulation across forethought, performance, and self-reflection, engaging in task analysis, simplifying language for audiences, continually evaluating clarity, and flexibly adapting across academic and science-communication genres with sustained attention to audience and purpose (Negretti et al., 2023).

The Present Study

This study addresses key gaps by (1) grounding AI feedback in the established genre-specific knowledge model countering the field's prevailing emphasis on lower-level issues at the expense of rhetorical appeal and revision strategies; (2) reducing variability by explicitly using the four domains to structure theory-grounded, holistic prompts and feedback; (3) incorporating rhetorical and process knowledge as core feedback dimensions and retaining teacher and peer feedback; (4) and, finally, using a natural-classroom design to avoid over-measurement of subjective consciousness; inductive analysis of in-situ reflections captures the lived experience of AI use.

Given these gaps, the research questions are:

1. How do genre-based ChatGPT feedback and interaction shape the basis, accuracy, and evolution of novice L2 writers' metacognitive judgments across Tardy's four domains of genre-specific knowledge—formal, rhetorical, process, subject-matter?
2. How do these judgments affect students' self-regulation strategies?

Methodology

Research Design, Setting, and Purpose

This study employs a qualitative, longitudinal study design to enable an in-depth exploration of L2 writers' metacognitive judgments and decision-making as they navigate AI-generated, genre-oriented feedback. The study was conducted in a first-year composition course for international students at a major U.S. university. The in-person course spanned one semester; students completed three major assignments across three genres: a literacy narrative, a genre analysis report, and a researched discourse community profile.

Participants and Sampling

Participants (ages 18–19 years) were international English-as-an-additional-language (EAL) students, enrolled in a first-year composition course at a U.S. university, all with at least intermediate L2 writing proficiency. In accordance with institutionally approved IRB protocols, 13 students provided informed consent. Purposeful sampling (Patton, 1990) then yielded a final sample of nine participants from the initial 13.

LLM Platform

The researcher created a custom GPT, *Genre Guru*, powered by GPT-4, which generates feedback across Tardy's four domains of genre-specific knowledge (Figure 3). Students did not have access to Genre Guru for the first third of the semester and used the free GPT-3.5 version. This logistical transition was treated as part of the natural ecology of the course and was incorporated into the analysis.

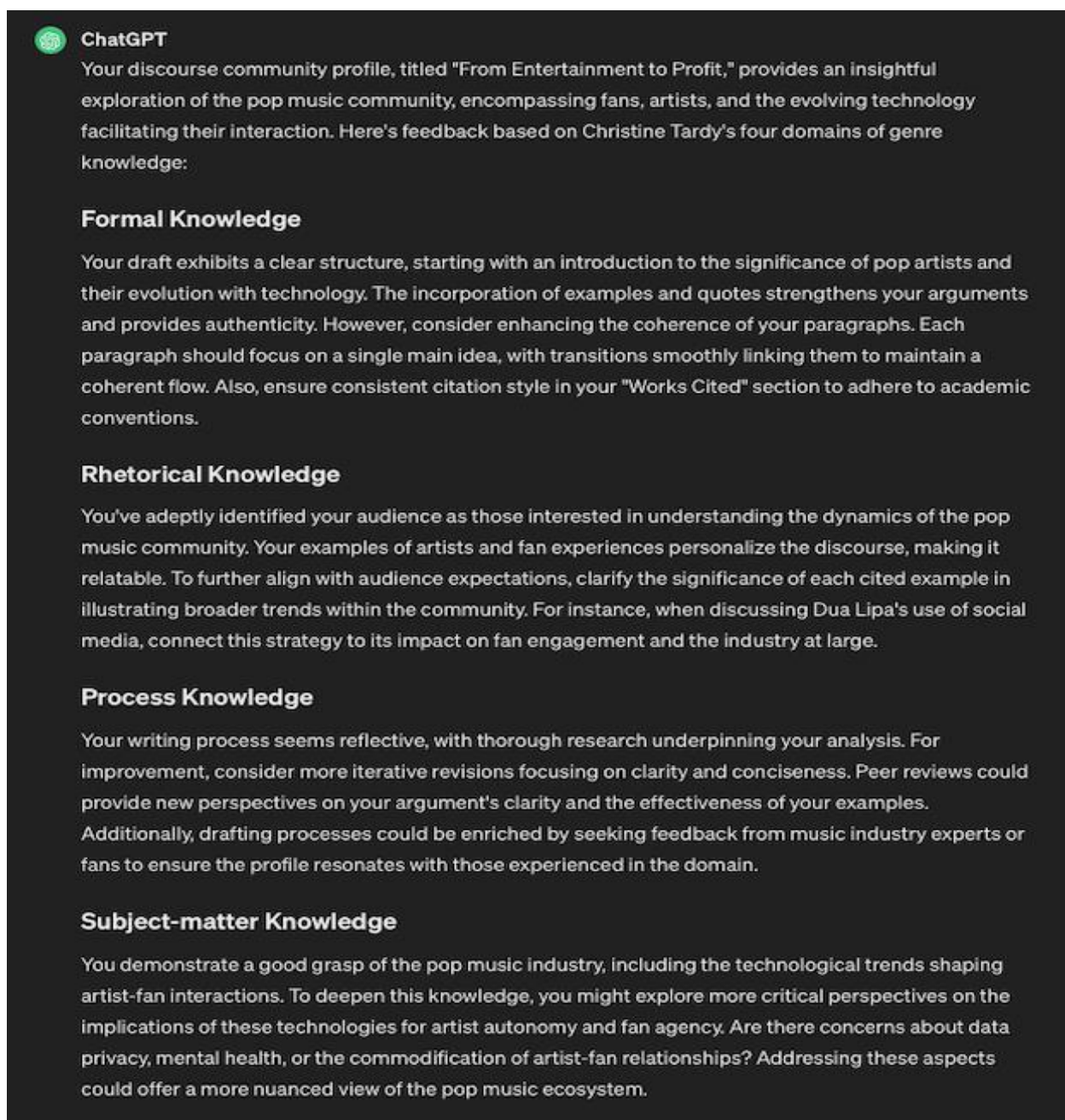


Fig. 3. Sample student feedback from ChatGPT framed via the four domains of genre-specific knowledge.

Procedure

For each major project, students produced a first and final draft. To capture immediate metacognitive judgments, students completed reflection prompts (see Appendix A), articulating

their experiences with ChatGPT and self-assessing their genre-specific knowledge across the four domains. The iterative procedure was as follows: students wrote a first draft; submitted it to ChatGPT for genre-based feedback; reflected on the feedback in writing; asked the AI at least 10 questions; then reevaluated and reflected on their genre-specific knowledge. Reflection logs, the primary data source, comprised two parts: a response immediately after receiving feedback and another after the interaction. This process was repeated across the three projects over the semester.

Data Collection

Multiple sources were collected and triangulated to construe a holistic understanding of L2 writers' metacognitive judgments and self-regulation strategies in response to ChatGPT feedback. The primary source was the AI genre-feedback reflections (after feedback and after interaction). ChatGPT interaction logs were also collected to examine questioning patterns and contextualize evolving judgments. At the end of the course, final AI reflections provided a global view of students' experiences. Finally, five participants completed one-hour, semi-structured interviews post-semester focusing on experiences with ChatGPT and perceived genre-specific knowledge.

Operational Definitions

Because the study focused on students' subjective assessment of their own learning and how those assessments evolve, essay drafts were not collected for analysis. Accuracy, usually gauged through performance, was instead operationalized as *calibration*, defined as the degree to which students' metacognitive judgments of their writing—across Tardy's four domains—match the demands of the writing tasks (e.g., task requirements, established genre conventions) and the feedback they receive (e.g., from ChatGPT, the instructor, or peers).

Data Analysis

Data were triangulated and analyzed qualitatively to provide a holistic view of students' MJs and self-regulation strategies, focusing on the basis, accuracy, and evolution of judgments and their relation to Zimmerman's three phases of self-regulation (forethought, performance, self-reflection). The AI Genre-feedback reflections, interaction logs, final AI reflections, and interviews were analyzed thematically (Braun & Clarke, 2006). This essentialist/realist reflexive thematic analysis was more theoretical than inductive in orientation, given the coding constraints imposed by the research questions, while still allowing flexible space for emerging themes. In alignment with the goal of understanding the rationale for L2 writers' judgments, the analysis was conducted at the latent level to consider underlying ideas and assumptions theorized as shaping the semantic content of the data (Braun & Clarke, 2006, p. 84).

Data analysis was conducted in NVivo. Each participant was treated as a finite case and coded longitudinally across course projects so that developmental changes could be traced before any cross-case analysis. All materials—AI genre-feedback reflection logs, final AI reflections, ChatGPT interaction logs, and interviews—were de-identified and organized by participant and project. Texts were segmented into meaning units that expressed a single metacognitive judgment (MJ) and/or the immediately ensuing self-regulation (SR) move. Each unit was tagged with participant, project, and source to maintain analytic traction on change over time.

Coding proceeded in two steps. First, metacognitive judgments based on metacognitive knowledge were deductively coded using Negretti's (2021) metacognitive knowledge categories: Genre Analysis, Reader Awareness, Rhetorical Strategies, and Writing Process. For every MJ, other emergent bases—the cues grounding the judgment (e.g., AI feedback, AI interaction, trust and skepticism, prior beliefs)—were additionally coded, and its calibration (whether the

judgment was warranted relative to genre demands and course criteria). The content of judgments was traced longitudinally to mark evolution, noting whether it shifted across projects. Self-regulation was coded only when it followed directly from a given MJ. Using Zimmerman and Moylan's (2009) model, I identified the phase implicated in the learner's next move—forethought, performance, or self-reflection. Operationally, “followed directly” referred to the next sentence/phrase or the same reflection entry. For instance, a statement like, “my rhetorical knowledge is still just around the beginning level; will try to engage with the audience by giving more solid examples that support my claims” was coded as MJ (basis: declarative metacognitive knowledge+audience engagement; calibrated; SRL: forethought (goal setting)).

In a second, inductive pass, data-driven subcodes were developed within and across these deductive families to capture recurring mechanisms—such as AI as a validator of writing practices and rhetorical intent, critical evaluation of AI feedback, confidence and self-efficacy, etc. Subcodes were iteratively refined through constant comparison, and the codebook was updated accordingly. Theme development followed the reflexive logic of moving from codes to patterned meaning. Subcodes were grouped into candidate themes, reviewed against coded extracts and the full dataset, and then refined and named, with careful attention to breadth across sources and timepoints. Findings are reported in the next section as thematic narratives that integrate analytic claims with excerpts. Each excerpt is situated within its metacognitive and, when relevant, regulatory context.

Findings

Theme 1: From Skepticism to Measured Trust in AI

Participants entered the semester with varying skepticism toward AI, which initially shaped their MJs and self-regulation. This skepticism formed the basis of participants'

metacognitive judgments. As they engaged more with the AI, these judgments shifted and, in turn, redirected their self-regulation.

At the outset, participants' prior beliefs shaped their MJs and tentative trust in AI. Some questioned AI's utility, relevance, and accuracy. Emir cast rhetorical awareness as beyond machines' grasp, stating, "I thought that ChatGPT or any other AI were not developed enough for that." These doubts underpinned participants' judgments, reflecting an initial belief that AI could not address higher-level genre knowledge.

Reluctance to accept AI advice prompted new self-regulation strategies to judge feedback and evaluate their own writing, marking the self-reflection phase. Skepticism yielded approaches such as cross-checking and partial acceptance, early markers of growing trust. They cross-checked AI comments against instructor guidance, which served as the benchmark for relevance and accuracy. When the two aligned, participants were more receptive to AI input. Caution also produced systematic accuracy checks, including intentionally steering the AI toward false or misleading answers, a planned forethought-phase tactic. Students used this especially to probe AI-flagged issues the instructor had not raised. Such deliberate testing illustrates how judgments evolved, with skepticism becoming proactive, investigative self-regulation that refines a conditional sense of ChatGPT's reliability.

With repeated AI interactions, participants' initial skepticism shifted to selective, measured trust in ChatGPT's feedback. Enough accurate feedback from ChatGPT challenged their beliefs. As Deniz noted: "Through the process my suspicions about ChatGPT accuracy got lost in the air, and I started to trust its answers, not believe but *trust* [emphasis added]." This shift illustrates how MJs moved from doubt to cautious acceptance, opening new self-regulatory receptivity to AI suggestions.

Perceived feedback quality also drove trust. Some students reported a clear improvement in AI answers during the second project's feedback round, when GPT-4 was introduced. This upgrade coincided with increased trust, vigilance, and critical engagement, prompting participants to reassess the AI's critiques. Moreover, participants valued ChatGPT's explanatory depth, which increased trust. For example, when Min-Jun asked for an evaluation of his genre-specific knowledge: "AI said my rhetorical knowledge is the best...and my worst part is formal knowledge. Then AI explained the reasons, which made me trust their feedback..." He internalized a sharper assessment of his declarative knowledge and demonstrated conditional knowledge by recognizing when and why to trust the feedback, which led him to prioritize revising formal over rhetorical knowledge—an inference characteristic of the self-reflection phase. He then added: "So, I will fix my formal knowledge first...instead of changing the rhetorical part of my report. I believe this process produces efficiency in my review." Here, Min-Jun's self-regulation follows from growing trust: he channels ChatGPT's advice into a new revision priority. The cycle closes with a clear causal attribution, signaling conditional and procedural metacognitive knowledge.

By the third assignment, participants' trust in AI was selective, not blind. For instance, Deniz maintained healthy skepticism in analyzing sources he had already used: "Genre Guru couldn't detect this because it does not build connections between citations or know their insights because of its own algorithm."

By course end, all participants expressed positive views of AI, seeing it as a supplement to learning, not a replacement. Their initial skepticism evolved into measured reliance, shaping self-regulatory writing that blended human judgment and AI input. This measured trust signaled skillful AI use without sacrificing creativity, originality, voice, or academic standards. Nor did

trust mean offloading writing's cognitive work to AI. As Participant 8 noted: "I don't think we should rely on it too much. If we continue to rely on AI to help us check or make suggestions for our articles, we will gradually lose our ability to think."

In short, skepticism prompted reflection on AI's role in learning. Skepticism became the basis for deeper reflection, while growing acceptance spurred new self-regulatory behaviors. It also helped participants self-assess the genre-specific knowledge domains they struggled with and devise strategies—through external or internal validation—to refine their MJs. This skepticism-regulation-reflection cycle sharpened judgments and increased confidence, cultivating measured trust that positioned AI as a complementary writing tool.

Theme 2: Critical Evaluation of AI Feedback and Ownership of the Writing Process

Participants critically examined the AI's suggestions. The synergy between evaluating feedback critically and integrating it with their own MK calibrated their MJs. Consequently, they adapted their SRL strategies to maintain ownership and agency over the writing process. From the beginning, participants' reasons for accepting or rejecting the feedback formed the basis for their MJs. They compared the AI's feedback with their own MK to inform those judgments. Rejection of the feedback—an adaptive self-reflection strategy—was often accompanied by causal attributions for justification. Participants rejected feedback when their knowledge of genre conventions contradicted the AI's or when the advice misaligned with task requirements. In the rhetorical domain, Deniz rejected ChatGPT's insistence on including counter arguments in his narrative: "I do not think this would be best...might damage literacy narrative's purpose."

Occasionally, participants reacted emotionally and impulsively rather than analyzing the feedback rationally. This emotional basis led to inaccurate MJs and hindered effective self-regulation. Participant 6 declared, "It was a bit upsetting...I did not write that. And I don't think

it will confuse the readers.” A closer look at the Chat log reveals she misread ChatGPT’s phrasing, which offered a suggestion she ultimately dismissed.

In the second project, students displayed stronger critical engagement. Participant 7 who had been dismissive of ChatGPT acknowledged that “...although it was wrong in some places, the advice...did have some merit. I have to look at the feedback objectively and critically”, reflecting conditional knowledge. Later on, he incorporated a suggestion from ChatGPT when the latter validated his vision and rhetorical intuition.

Participants were more likely to overcome AI-induced negative emotions when the AI validated their progress. Recognizing the AI’s nature was also a factor. Deniz appreciated the contrast between LLMs and humans; he credits ChatGPT’s “emotionless” algorithm for his own successful emotional regulation.

For the first half of the semester, students either overcorrected with excessive skepticism or underestimated the AI’s capabilities. By semester’s end, MJs became more calibrated, improving students’ selectiveness. Participant 6 recognized her rhetorical and formal missteps and devised a thesis-check strategy forecasting an entire SRL cycle:

“In the future, I will set a clear detailed thesis...And I would ask myself if points that I’m going to write in my body paragraphs directly related to the thesis or if they affect the purpose of my writing. If not, more research is needed.”

Participant 6 did not blindly accept the feedback; she internalized ChatGPT’s comments and devised her own revision plan: a planned strategy (forethought), monitoring it through questioning (performance), and evaluating its success via an if/then scenario (self-reflection).

At this stage, participants' rationale for rejecting the feedback was deeper and more nuanced, amplifying ownership. Deniz disputed ChatGPT's praise of his process knowledge because he felt his distributions of sources was unbalanced (contra ChatGPT's evaluation). He then set a new forethought-phase goal of integrating more academic sources and distributing them more evenly.

Rejection was not always a response to initial feedback; it emerged as students monitored their writing via interaction, where they controlled their performance. Diego actively declined to incorporate the AI's rhetorical strategy of comparing and contrasting pop song lyrics and rhymes, favoring his conditional knowledge: "...because the focus of this analysis was not about a musical genre but the writing genre of pop song lyrics...(it) is better not to add that point...that is not the intention." More generally, participants remained the sole arbiters of revision, integrating various sources of feedback through their own judgment and preserving agency.

The initial basis for critical evaluation evolved into a more sophisticated stance of selective acceptance and rejection, culminating in robust self-regulatory routines that emphasized students' ownership of writing. The interaction with the AI itself scaffolds one's MK, leading to newer MJs: "With each interaction, ChatGPT's responses and independent thinking together formed my understanding of writing and gave me new ideas for improvement" (Participant 7).

Students' critical use of AI was not just a temporary educational necessity; it was essential for building transferrable skills beyond the classroom. All participants acknowledged the AI's power to supplement their learning but firmly opposed over-reliance on the technology.

Ultimately, the ongoing interaction with ChatGPT brought about a change in mindset: students over time acknowledged the AI's power to augment their learning, but firmly maintained their role as final decision-makers. The combination of openness, skepticism, and

awareness fostered effective self-regulation, ensuring that the technology is used responsibly. Participants also shared a deeper insight: while ChatGPT can close their metacognitive gaps, they themselves control how the AI is harnessed. By relying on their own MJs, they ensured that the integration of AI-feedback was mindful, strategic, and critical.

Theme 3: Holistic Integration of Tardy's Four Domains and The Emergence of Genre Awareness

Before receiving AI feedback, students based their MJs on their confidence level and antecedent MK. ChatGPT's feedback, however, became the primary basis for their judgments recalibrating assessments. Participants were initially overconfident in their formal knowledge. Yet the realization that formal knowledge was weaker than assumed directly prompted new self-regulatory actions, including consistent proofreading and double-checking spelling. Participants also relied on their antecedent MK and prior experiences when gauging their understanding, but the AI feedback became the new basis for judgments, replacing prior beliefs. Participant 8 realized the importance of genre conventions, prompting her to recalibrate: "ChatGPT kept reminding me to add personal experiences and examples. This made me realize that literacy narrative needs vivid, natural examples and personal growth...I will add more details on the processing of my growth..." This shift demonstrates how revised MJs about what constitutes a strong literacy narrative led directly to new writing behaviors.

Participants' rhetorical knowledge was nascent early on partly because they mostly wrote for themselves. After ChatGPT's feedback, students became more aware of audience considerations. By the second project, students adopted a more audience-centric approach to writing. They became increasingly aware of how their subject-matter affected audience awareness. Interaction with ChatGPT helped Participant 6 refine her audience. The feedback she

co-created with the AI triggered a shift in her declarative and conditional knowledge, leading to a chain of causal attribution: “I used to think that it would be easier to describe the connection between the adults and the short stories...I was wrong...Adults play a significant role in delivering the message...they should not be only mentioned as a secondary audience...” The cause-and-effect sequence is clear: once she recognized a gap in her rhetorical assumptions (a MJ), she specifically targeted the mismatch, exemplifying how metacognitive evolution directly yields self-regulation.

ChatGPT’s feedback allowed students to test whether their purpose was clear, thus serving as a validator of rhetorical intent. When ChatGPT’s response conflicted with the writer’s intended purpose, writers’ reconsidered their main claims and thesis statement in the next forethought phase. While feedback and interaction contributed to forming judgments, it was conditional and procedural metacognitive knowledge that played the most significant role. Li-Hao became more aware of the importance of manipulating the subject-matter for rhetorical effects: “my subject-matter knowledge has shifted...to in-depth knowledge by providing more examples about specific topics using a third-person perspective...I may add some scholarly papers as an example of explaining a complicated subject...”

Participants proactively sought ChatGPT’s assessment of their genre-specific knowledge. They either asked for a ranking, the best/worst domain, or a numerical grade. This approach allowed them to monitor their progress and ensure their self-regulation targeted their weakest domains. By the end of the semester, participants’ MJs became more aligned with the AI’s; consequently, their self-regulation strategies increasingly favored holistic and balanced development of the four domains: “I am not going to change a lot because I should improve my other knowledges to balance my writing” (Min-Jun).

The feedback loop with the AI across multiple projects promoted genre awareness. Emir demonstrated a higher-order grasp of genre awareness by providing a rare, explicit meta-reflection on how the domains intersect:

“Before the feedback loop, I was thinking about four domains of genre knowledge as a more separate thing. However...I learned that they are interconnected with each other...without rhetorical knowledge about the genre determining the formal knowledge is really hard... the formal and rhetorical knowledge intercepted with each other and helped each other develop.”

Emir’s realization suggests that engagement with AI feedback can foster a deeper reflection on how they intertwine, offering the clearest evidence that ChatGPT interaction can culminate in an explicit, theoretical form of genre awareness.

ChatGPT served both as a mirror, revealing hidden gaps, and as a catalyst for refining judgments, culminating in deliberate and targeted self-regulation strategies, such as systematic proofreading, clarifying rhetorical goals, or selectively deepening subject content. This process expedited students’ integration of the four domains and their mastery of the writing process.

Theme 4: Emotions and Motivation as Drivers of Self-regulation

Emotions and motivation underpinned students’ MJs. They shaped participants’ reassessment of Tardy’s four domains, altering their self-beliefs and perceived MK gaps. With AI use, students’ emotions shifted from negative to positive. In turn, their confidence and self-efficacy grew, accelerating their SRL cycles.

Initial feedback surfaced strengths and weaknesses, fostering calibrated confidence. Some reported low confidence, rooted in prior declarative metacognitive beliefs. ChatGPT's feedback challenged these assumptions, calibrating their MJs and confidence. Conversely, some were overconfident. That overconfidence came with surprise, which mediated MJ recalibration. This surprise centered on subject-matter knowledge. Students underestimated elaboration, exemplification, and content selection. Participant 9 noted: "I was just surprised on how it still talks about elaborating more on some topics in which I felt I had elaborated enough."

Surprise signaled a mismatch between participants' beliefs about AI and its observed performance. When AI and instructor feedback aligned, Emir—initially in disbelief—reported greater trust in ChatGPT. Participant 8 reported similar surprise when the AI partially validated her draft.

The AI's non-human status reduced feedback anxiety, supporting self-regulation: "it's first-time listening feedback from someone/something who is not professor or instructor which makes process more relief if you are afraid of being criticized. I hope to...apply this experience to other genres" (Deniz). This reduced anxiety made participants more receptive to critical suggestions, further facilitating self-regulation.

Participants typically reported positive emotions such as joy and satisfaction when ChatGPT validated their effort and domain-specific gains. These emotions reinforced self-reflection, motivating refinements to genre knowledge in the performance phase. This validation built confidence, yielding more ambitious revisions. Diego noted:

“I am really happy to see that my genre analysis and report are well-focused, now I have to arrange those details, as well as add some information from the direct quotation that would be helpful in the understanding of the genre.”

This reinforcement increased confidence and strengthened perceived genre competence, prompting bolder SRL goals. By affirming Diego’s rhetorical intent and content coverage, ChatGPT’s feedback strengthened his self-belief, initiating a new SRL cycle.

ChatGPT’s validation turned negative reactions positive and anchored a realistic appraisal of strengths and weaknesses, boosting self-efficacy. ChatGPT’s validation of Participant 7’s audience awareness reinforced his self-efficacy: “...I know who my audience is. My topic is the Steam community...So my hands are almost full of people who are interested in video games and...people who are interested in digital networks.” This explicit approval shows how the positive emotional shift prompted Participant 7 to reaffirm his rhetorical knowledge, strengthening his SRL strategy. Heightened self-efficacy matters because it reaffirms students’ agency and autonomy over the writing process, which AI can sometimes undermine.

Self-efficacy catalyzed resilience and a growth mindset. Li-Hao felt “deeply depressed” after reading AI feedback. His reaction reflected a mismatch between his performance and perceived skill. Although AI did not fully validate his formal knowledge, it affirmed his organizational skills. That praise—recognizing structural work begun in a prior writing course—shifted his stance, fostering resilience, self-efficacy, and a growth mindset, initiating a new forethought phase: “I believe instead of complaining and criticizing myself, efficiently adjusting my weaknesses can significantly improve my writing.”

In short, ChatGPT's partial validation of specific strengths helped participants recalibrate their MJs and persist, showing how AI feedback shapes writers' self-assessment and subsequent self-regulation. By semester's end, students' confidence and motivation peaked. Exhilarated by a major performance gain, Li-Hao reported a shift in writer identity, adopting a more adaptive, resilient stance that supported self-regulation:

“My heart was bumping like receiving a 90% grade... The first word that came to my mind was, ‘WOW, I achieved it.’ ... I have often been called a ‘Bad Writer’ for so long...The huge encouragement from ChatGPT inspired me to continue to write and adjust my weaknesses.”

Surprise and other emotional reactions diminished as students' MJs aligned more closely with AI judgments. In turn, their developing growth mindset prompted adoption of new SRL strategies and bolstered confidence.

Deniz noted that proactive engagement with AI demystified the writing process, making him more willing to embrace criticism: “I feel more confident...because even if I do a mistake... I could put it into ChatGPT and get the right feedback. It helped me understand that making mistakes is possible, but correcting them is also possible.”

In sum, engagement with ChatGPT recalibrated the emotional basis of MJs. Over time, emotions more accurately reflected expertise. As confidence and self-efficacy—guided by partial AI validation—grew, revision practices improved. In turn, updated SRL strategies improved performance and emotions, fueling persistence and adaptability and reinforcing a growth mindset.

Discussion

This study examined students' decision-making processes as they navigate ChatGPT's genre-focused corrective feedback. Drawing on the frameworks of metacognitive judgments (Koriat, 1997; Dunlosky & Thiede, 2013) and self-regulation (Zimmerman, 2000; Zimmerman & Moylan, 2009), this study focused on the cognitive and behavioral aspects of learning to write in a second language, specifically, how novice L2 writers self-assess their learning through interactions with AI and, in turn, translate these assessments into concrete self-regulation strategies. Accordingly, the study addressed two research questions. The first investigated how ChatGPT's feedback and interactions influence three features of MJs—basis, accuracy, and evolution—across Tardy's four domains of genre-specific knowledge. The second explored how these judgments affect students' self-regulation strategies.

In summary, four themes explain the dynamics between students' metacognitive judgments and self-regulatory behaviors. First, participants moved from initial skepticism to measured trust in ChatGPT's feedback. Second, students critically evaluated ChatGPT's suggestions while maintaining ownership of their writing. Third, over time, novice L2 writers integrated the four domains of genre-specific knowledge (formal, rhetorical, process, subject-matter), fostering genre awareness. This pattern echoes prior findings that explicit metacognitive work can accelerate domain integration (Negretti & McGrath, 2018; Negretti, 2021). Fourth, students' emotions and motivation were primary drivers of their self-regulation cycles. Due to students achieving higher metacognitive and motivational via AI interaction, these findings suggest that LLMs in L2 writing instruction ought to be integrated by default with explicit safeguards and opt-out alternatives.

For RQ1, ChatGPT's initial feedback and subsequent dialogue were the primary catalysts in calibrating participants' metacognitive judgments over time across the formal, rhetorical, process, and subject-matter domains. Feedback from ChatGPT initially sparked skepticism among participants, who doubted AI's potential to supplement their learning. However, this skepticism evolved into measured trust as students triangulated the feedback with other sources and came to appreciate the technology's affordances. This trajectory toward measured trust mirrors prior AWE research (Ranalli, 2021) and more recent work indicating that trust in LLMs grows with perceived model competence (Bouyzourn & Birch, 2025). The shift from skepticism to measured trust was important because it established AI feedback as credible, reliable, and worthy of examination. Once a baseline of trust had been established, students not only examined ChatGPT's suggestions critically but also strengthened their critical thinking to evaluate the feedback, indicating their evolving metacognitive awareness. Thus, critical evaluation emerged as both a process and an outcome that students deliberately pursued.

Regarding the development of trust, instructors can invite students to reflect on or annotate AI feedback and comment on whether they trust the suggestions at hand and why. Critical evaluation of AI output emerges as a fundamental skill for navigating AI use. Hence, critical AI literacy should be taught explicitly as a metacognitive strategy. Instructors can facilitate this task by designing and formalizing prompts that push students to question the feedback or even demand that the LLM justify its responses with a rationale. To mitigate risks (e.g., hallucinations, over-reliance), instructors can require sources for factual claims, voice-preservation prompts, and limit the cap permissible for AI interventions per draft.

AI feedback and interactions also anchored students' metacognitive judgments in internal criteria—namely, their metacognitive knowledge. As participants became increasingly adept at

balancing ChatGPT's suggestions with external feedback, they turned their attention inward to reflect on their declarative, procedural, and conditional metacognitive knowledge. This shift unfolded in parallel with growing trust as ChatGPT gained credibility in students' eyes. Consequently, participants moved from a stance-based adoption of feedback to an evidence-based approach: they were able to formulate rationales for including specific suggestions and estimate the feasibility of implementing them, thereby honing their conditional and procedural knowledge. In doing so, they affirmed ownership of their writing by making final decisions based on a holistic understanding of the writing situation.

With trust and critical thinking honed, participants recognized the interconnectedness of the four domains of genre-specific knowledge. Students readily integrated formal and subject-matter feedback but initially struggled with the rhetorical and process domains. This trajectory is unsurprising, given the dominance of form-focused instruction in English language teaching. Most importantly, the findings of this study suggest that feedback structured around the four domains of genre-specific knowledge can serve as the backbone for L2 writing instruction. In practice, instructors can lead students in metacognitive workshops around the four domains, thus providing common vocabulary they can rely on for the rest of the course. Assignments and rubrics need to be realigned to reflect the focus on genre-specific knowledge with the four headings made explicit.

A major finding is that participants showed high plasticity in assimilating feedback on rhetorical and process knowledge—a trait typically observed in more advanced L2 writers (e.g., Negretti, 2017). In the rhetorical realm, all participants explicitly linked form and subject matter to audience conceptualization and engagement. Viewing their writing from the reader's perspective proved pivotal for developing students' internal models of the communicative

situation and equipping them to make informed choices across contexts. As for process knowledge, students became metacognitively aware of the steps necessary to produce genres, including drafting and revising, researching and integrating sources, and verifying subject-matter accuracy. Because rhetorical and process knowledge were slower to develop, instructors should prioritize targeting those weaker domains with rigorous instruction in the initial phases of the course then rebalance as calibration improves.

More importantly, several participants recognized that written reflection in response to AI feedback served as a procedural scaffold in itself. The reflection logs used in this study were not merely data-collection tools; they functioned as channels for learning transfer across assignments—or, in Deniz’s words, “a bridge between projects.” As was implemented in this study, instructors might envisage embedding reflection loops to drive self-regulation. Where feasible, reflections can be paired with brief artifact uploads (e.g., annotated screenshots of accepted/rejected AI suggestions) to make reasoning visible and gradable. Moreover, sample excerpts from prior students can serve as a model for students to witness high quality metacognitive reasoning.

Students’ adaptability to the feedback stems from the four-domain categorization and the theory’s robustness and coherence, which, when scaffolded through AI interactions, can demystify scholarly concepts for students in real time. This suggests that reverse-engineering theoretical models with LLMs is not only desirable for promoting learning but also necessary to test and refine existing theory. Moreover, students’ adaptability in this study suggests that novice L2 writers possess an underlying cognitive apparatus beyond traditional language proficiency that can be activated through AI dialogue, lending further support to language-independent

processes (Tardy et al., 2020), such as genre awareness—an emergent quality particularly pronounced for Deniz, Emir, and Diego.

As for RQ2, students' MJs determined which self-regulatory pathways they pursued. Regarding accuracy, although their initial MJs were understandably imprecise—given that they were taking the course to become better writers—their openness to feedback enabled them to update their metacognitive models gradually through repeated reflection–forethought cycles, marked by adaptive inferences often translated into future goals. Students' emotional reactions, such as surprise, were crucial: they alerted them to potential knowledge gaps. Relief was another emotion that surfaced whenever ChatGPT confirmed writers' existing moves, an emotion that appeared to accelerate commitment to subsequent revisions. More generally, explicit praise from ChatGPT often triggered positive affect, which accelerated self-efficacy gains. Pedagogically, LLMs can be used as an affective buffer that boosts self-confidence and seeds a growth mindset. International students, who might experience language anxiety in a foreign country, might benefit from AI feedback early in the course to normalize feedback reception, provide a safe space for emotional exploration, and affectively prime them for teacher feedback.

The basis of students' MJs significantly affected the pace of self-regulation. Students adopted ChatGPT's suggestions almost automatically when the instructor corroborated them or when they found the rationale convincing. By contrast, self-regulation slowed when writers evaluated novel recommendations against their metacognitive knowledge, prompting deeper critical thinking. In such cases, students often articulated additional causal attributions before deciding on their next step—adopt the feedback fully or partially, or reject it altogether.

More generally, students' metacognitive judgments gained motivational resonance over time, becoming imbued with self-efficacy and perseverance and culminating in a growth

mindset—factors that collectively accelerated repeated self-regulation cycles. This finding aligns with recent scholarship showing a positive impact of ChatGPT on motivation and overall well-being (Wang et al., 2025). A key mechanism underlying this transition was ChatGPT’s role as a validator of students’ writing practices and rhetorical intent. This validation was not only emotionally satisfying but also motivational. In other words, validation permeated both the self-reflection and forethought phases, affirming what students had done and could do. Moreover, as students deepened their grasp of the four domains, their confidence increased, leading to stronger self-efficacy. Crucially, students’ perception of writing shifted, through interactions with ChatGPT, from viewing writing as a fixed ability to believing that mastery is attainable with sufficient dedication. This growth mindset informed their subsequent interactions with AI and persisted beyond the course.

Conclusion

This study examined L2 writers’ metacognitive processes in a context where LLMs provide genre-based feedback. When scaffolded by Tardy’s four-domain framework, ChatGPT reprioritized the basis of writers’ MJs, improved their calibration to genre demands and the feedback ecology, and supported the evolution of those judgments across the semester. Decisions increasingly drew on declarative, procedural, and conditional knowledge, strengthening ownership of the writing process. Self-regulation accelerated only after measured trust in the tool emerged; once established, writers made more adaptive inferences, initiated forethought, monitored performance through interaction, and closed the loop with self-reflection. These gains were amplified by self-efficacy, persistence, and a growth mindset—partly catalyzed by ChatGPT’s role as a validator of writing practices and rhetorical intent.

Nonetheless, several limitations qualify these claims and motivate future work. The multiple-case design privileged depth over breadth, constraining transferability. The naturalistic model upgrade constitutes a potential confound for interpreting developmental change. Focusing on internal metacognitive processes limits causal claims about product-level gains, and reflections, serving as both scaffold and data source, introduce measurement reactivity. Finally, the instructor-researcher role entails familiar reflexivity concerns. The next wave of studies can (a) broaden contexts and participants, (b) stabilize the AI, (c) pair mechanism-rich traces with objective outcomes and calibration indices, and (d) reduce measurement reactivity with fine-grained process-tracing (e.g. screen capture). These steps will help determine the extent to which the trajectories observed here generalize beyond this single setting and translate into durable gains in L2 writing. In short, while LLMs do not possess subjective experience, they can scaffold genre development and metacognitive decisions that L2 writers leverage to regulate their writing; used this way, AI-calibrated metacognition emerges from treating LLMs as a metacognitive catalyst—not a replacement for human judgment.

References

- Aladini, A., Ismail, S. M., Ahmad Saleem Khasawneh, M., & Shakibaei, G. (2025). Self-directed writing development across computer/AI-based tasks: Unraveling the traces on l2 writing outcomes, growth mindfulness, and grammatical knowledge. *Computers in Human Behavior Reports*, 17, 100566. <https://doi.org/10.1016/j.chbr.2024.100566>
- Apriani, E., Cardoso, L., Obaid, A. J., Muthmainnah, Wijayanti, E., Esmianti, F., & Supardan, D. (2024). Impact of ai-powered chatbots on efl students' writing skills, self-efficacy, and self-regulation: A mixed-methods study. *Global Educational Research Review*, 1(2), 57–72. <https://doi.org/10.71380/gerr-08-2024-8>
- Bouyzourn, K., & Birch, A. (2025). *What shapes user trust in chatgpt? a mixed-methods study of user attributes, trust dimensions, task context, and societal perceptions among university students*. arXiv.org. <https://doi.org/10.48550/arXiv.2507.05046>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chen, Z., Wei, W., Zhu, X., & Yao, Y. (2025). Unpacking the rejection of l2 students toward chatgpt-generated feedback: An explanatory research. *ECNU Review of Education*. <https://doi.org/10.1177/20965311241305140>
- Driscoll, D. L., Paszek, J., Gorzelsky, G., Hayes, C. L., & Jones, E. (2019). Genre Knowledge and Writing Development: Results From the Writing Transfer Project. *Written Communication*, 37(1), 69–103.
- Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. SAGE Publications .

- Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, 24, 58–61. <https://doi.org/10.1016/j.learninstruc.2012.05.002>
- Ellis, R. (2010). Epilogue. *Studies in Second Language Acquisition*, 32(2), 335–349. <https://doi.org/10.1017/s0272263109990544>
- Fathi, J., & Rahimi, M. (2024). Utilising artificial intelligence-enhanced writing mediation to develop academic writing skills in efl learners: A qualitative study. *Computer Assisted Language Learning*, 1–40. <https://doi.org/10.1080/09588221.2024.2374772>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34(10), 906.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. *Perspectives on writing: Research, theory, and practice*, 6–44.
- Jin, F., Lin, C.-H., & Lai, C. (2025). Modeling ai-assisted writing: How self-regulated learning influences writing outcomes. *Computers in Human Behavior*, 165, 108538. <https://doi.org/10.1016/j.chb.2024.108538>
- Kamelia, L., & Linda, C. (2023). *An investigation of the role of artificial intelligence in promoting EFL learners' self-regulated learning skills: The case of master's students at the department of English at MMUTO* [Doctoral dissertation, Mouloud Mammeri University].
- Koltovskaia, S., Rahmati, P., & Saeli, H. (2024). Graduate students' use of chatgpt for academic text revision: Behavioral, cognitive, and affective engagement. *Journal of Second Language Writing*, 65, 101130. <https://doi.org/10.1016/j.jslw.2024.101130>

- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349–370. <https://doi.org/10.1037/0096-3445.126.4.349>
- Lee, Y.-J. (2024). Can my writing be polished further? when chatgpt meets human touch. *ELT Journal*, 78(4), 401–413. <https://doi.org/10.1093/elt/ccae039>
- Li, J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18.
- Ma, X., Zhang, C., & Lee, I. (2025). *Learner engagement with ai-generated feedback among chinese efl students on second language writing*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5197066
- Mohammed, S., & Khalid, M. (2025). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia*, 15(1), 7. <https://doi.org/10.1186/s40468-025-00343-2>
- Negretti, R. (2012). Metacognition in student academic writing: A longitudinal study of metacognitive Awareness and Its relation to task perception, self-Regulation, and evaluation of performance. *Written Communication*, 29(2), 142–179.
- Negretti, R. (2017). Calibrating genre: Metacognitive judgments and rhetorical effectiveness in academic writing by L2 graduate students. *Applied Linguistics*, 38(4), 512–539.
- Negretti, R. (2021). Searching for metacognitive generalities: Areas of convergence in learning to write for publication across doctoral students in science and engineering. *Written Communication*, 38(2), 167–207. <https://doi.org/10.1177/0741088320984796>

- Negretti, R., & Kuteeva, M. (2011). Fostering metacognitive genre awareness in L2 academic reading and writing: A case study of pre-service English teachers. *Journal of Second Language Writing*, 20(2), 95–110. <https://doi.org/10.1016/j.jslw.2011.02.002>
- Negretti, R., & McGrath, L. (2018). Scaffolding genre knowledge and metacognition: Insights from an L2 doctoral research writing course. *Journal of Second Language Writing*, 40, 12–31.
- Negretti, R., Sjöberg-Hawke, C., Persson, M., & Cervin-Ellqvist, M. (2023). Thinking outside the box: Senior scientists' metacognitive strategy knowledge and self-regulation of writing for science communication. *Journal of Writing Research*, 15(2), 333–361. <https://doi.org/10.17239/jowr-2023.15.02.04>
- Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, 46(5), 756–778. <https://doi.org/10.1080/02602938.2020.1823314>
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). SAGE Publications, Inc.
- Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing*, 52, 100816. <https://doi.org/10.1016/j.jslw.2021.100816>
- Shen, X., & Tao, Y. (2025). Metacognitive strategies, AI-based writing self-efficacy and writing anxiety in AI-assisted writing contexts: A structural equation modeling analysis. *International Journal of TESOL Studies*, 7(1), 70–87. <https://doi.org/10.58304/ijts.20250105>

- Shraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary educational psychology*, 19(4), 460–475.
- Tam, A. (2025). Interacting with chatgpt for internal feedback and factors affecting feedback quality. *Assessment & Evaluation in Higher Education*, 50(2), 219–235.
<https://doi.org/10.1080/02602938.2024.2374485>
- Tardy, C. M. (2009). *Building genre knowledge*. Parlor Press.
- Tardy, C. M., Sommer-Farias, B., & Gevers, J. (2020). Teaching and researching genre knowledge: Toward an enhanced theoretical framework. *Written Communication*, 37(3), 287–321. <https://doi.org/10.1177/0741088320916554>
- Teng, M. (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>
- Wang, C., Wang, B., & Xu, D. (2025). The role of chatgpt and grammarly in promoting emotion regulation, psychological well-being, motivation, and academic writing in Chinese College students: A self-determination theory perspective. *Learning and Motivation*, 90, 102131. <https://doi.org/10.1016/j.lmot.2025.102131>
- Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, 3(1), 22–36.
<https://doi.org/10.1080/15544800701771580>
- Yan, D. (2024). Feedback seeking abilities of l2 writers using chatgpt: A mixed method multiple case study. *Kybernetes*, 54(7), 3757–3781. <https://doi.org/10.1108/k-09-2023-1933>

- Yan, D., & Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by chatgpt: A mixed-method multiple case study. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03543-y>
- Yao, Y., Sun, Y., Zhu, S., & Zhu, X. (2024). A qualitative inquiry into metacognitive strategies of postgraduate students in employing chatgpt for english academic writing. *European Journal of Education*, 60(1). <https://doi.org/10.1111/ejed.12824>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/b978-012109890-2/50031-7>
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299–315). Routledge. <https://doi.org/10.4324/9780203876428>
- Ziqi, C., Xinhua, Z., Qi, L., & Wei, W. (2024). L2 students' barriers in engaging with form and content-focused ai-generated feedback in revising their compositions. *Computer Assisted Language Learning*, 1–21. <https://doi.org/10.1080/09588221.2024.2422478>

Appendix A

AI Genre-feedback Reflection Prompts

Before completing this assignment make sure you have purchased the GPT-4 subscription and that you are logged into your GPT-4 account

Part 1: Initial Feedback

Step 1:

Log into your GPT-4 account and access the Genre Guru. This custom GPT will give you genre-specific feedback. Ask it to provide you with feedback by clicking on “Provide feedback on my genre analysis report and personal profile report”. Genre Guru will ask you to enter your draft. Copy and paste your first draft and click enter. Genre Guru will provide you with feedback on the four genre-knowledge domains.

Step 2:

Reflect on the feedback provided by ChatGPT in the form of a journal entry. Be as detailed and introspective as possible. Answer the following questions:

- a. Perception and Understanding: How did the feedback change your view of your writing skills? What new insights did you gain?
- b. Assessment of Genre-Specific Knowledge: Evaluate your application of genre-specific knowledge. Were there areas where you excelled or needed improvement in the formal, rhetorical, process, and subject-matter domains? In what areas is your genre-specific knowledge strongest? how do you evaluate your understanding and application of your own genre-specific knowledge across the four categories listed on the feedback? Elaborate and provide examples. Be detailed and thorough here.

c. Initial Reaction: Reflect on your initial thoughts upon receiving the feedback.

What were your first thoughts about the feedback? Which parts resonated with you, and were there any surprises? Elaborate on these points.

Before moving to the next part, ensure you have already answered all questions in part 1.

Part 2: Iterative Feedback

Step 1: Engage in Dialogue

Continue your discussion with Genre Guru, asking follow-up questions or seeking clarification to deepen your understanding. Aim for at least 10 interactions to explore various facets of your writing and the feedback received.

Step 2: repeat from Part 1 above.

Step 3: Please copy and paste a link to your ChatGPT dialogue log as part of your submission.

Appendix B

Table 1. Cross-case coding themes

Global/Longitudinal Themes	Organizing Themes	Initial themes
1. From Skepticism to Measured Trust in AI	1.1. Convergence/triangulation builds trust	Trust and Skepticism AI as Validator of Knowledge, Writing Practices Integration of Multiple Feedback Sources
	1.2 Reliability testing & appropriate use	Strategic Interaction and questioning Developing Critical Thinking through AI Critical Thinking and Feedback Evaluation
	1.3 Explanatory depth leads to prioritization	Critical Thinking and Feedback Evaluation AI as a Collaborative Partner
	1.4 Ethical boundary-setting	Trust and Skepticism AI as Validator.
	1.5 Perceived model competence & vigilance	

2. Critical Evaluation of AI Feedback & Ownership of the Writing Process	2.1 Evidence-seeking acceptance/rejection	Critical Evaluation of AI Feedback Critical Thinking and Feedback Evaluation
	2.2 Validation of rhetorical intent as catalyst	AI as Validator of rhetorical intent Understanding & Engaging Audiences
	2.3 Co-creation & prompt steering	Strategic Interaction and questioning Developing Critical Thinking through AI AI as a Collaborative Partner
	2.4 Workflow reconfiguration (ownership)	Changes in writing approach due to AI
	2.5 Atypical cases	Emotional Engagement and Maturation with AI Feedback
3. Holistic Integration of Tardy's Four Domains & Emergence of Genre Awareness	3.1 Reader-oriented rhetorical control	Audience Awareness & Rhetorical Engagement Understanding & Engaging Audiences Creating Arguments, Persuasion, and Narrative
	3.2 Form in service of rhetoric	Formal Knowledge Mastery for Persuasion and Audience Genre Conventions and Formal Elements
	3.3 Coordination & transfer across domains	Interconnectedness of the genre-specific domains

		Genre variation
	3.4 Monitoring via AI grading/ranking	Strategic Interaction and questioning Critical Thinking and Feedback Evaluation
	3.5 Contextual constraints reframed as resources	Language, Education, and Cultural Barrier Awareness of current practices and language choices
4. Emotions & Motivation as Drivers of Self-Regulation	4.1 Confidence calibration via partial validation	Confidence and Self-efficacy AI as Validator of Writing Practices
	4.2 Non-human feedback lowers anxiety; surprise fades	Emotional Engagement and Maturation with AI Feedback Trust and Skepticism
	4.3 Growth mindset, resilience, persistence	Resilience, Motivation, and Adaptability