Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 6, 43-54
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i6.7751

Generative AI–mediated scaffolds for enhanced critical thinking in EFL writing

Hui Hong[1*], Poonsri Vate-U-Lan[2], Chantana Viriyavejakul[3]
[1,2,3]School of Industrial Education and Technology, King Mongkut's Institute of Technology Ladkrabang, Thailand; 63603167@kmitl.ac.th (H.H.).

Abstract: Generative AI tools present new opportunities for enhancing critical thinking (CT) in English as a Foreign Language (EFL) writing instruction. This study investigates how the structured integration of these technologies could support the development of CT skills, particularly in vocational education contexts. An eight-week multiple-case action research design was conducted across three vocational colleges, involving 92 students engaged in iterative writing and revision cycles guided by the GenAI-CT framework. This pedagogical model draws on Bloom's taxonomy, Vygotsky's Zone of Proximal Development, and cognitive apprenticeship theory to scaffold learners through increasingly complex reasoning tasks. Data were collected from student essays, AI interaction logs, reflective journals, and classroom observations. Mixed-methods analysis revealed statistically significant gains across all CT dimensions ($p < .001$). Thematic findings indicated notable increases in analytical depth, metacognitive reflection, and evaluative judgment. Variations across cases underscored the influence of disciplinary focus and instructional mediation styles. These results demonstrate that generative AI, when embedded in intentional pedagogical structures, can foster cognitive engagement rather than superficial automation. The GenAI-CT framework offers a replicable model for integrating AI in applied language and communication instruction, supporting educators in cultivating critical thinking through technology-enhanced learning environments.

Keywords: *AI-Mediated learning, Cognitive scaffolding, Critical thinking, EFL writing instruction, Generative AI.*

1. Introduction

Over the past decade, English as a Foreign Language (EFL) writing instruction has faced two intertwined challenges. First, learners often produce grammatically correct but formulaic texts, lacking the depth of argumentation and independent reasoning prized in academic and professional contexts [1]. Second, while critical thinking is universally acknowledged as a key 21st-century skill, EFL curricula frequently emphasize discrete language forms over higher-order cognitive processes, leaving learners underprepared to analyze, evaluate, and generate nuanced arguments in English [2].

Generative artificial intelligence (GenAI) tools can produce draft text, suggest stylistic revisions, and generate reflective prompts that, if properly harnessed, may move learners beyond surface-level corrections toward deeper engagement with ideas [3]. Yet unstructured use of generative AI also risks fostering dependency, whereby students accept AI suggestions uncritically and miss opportunities to exercise their own analytical and evaluative capacities [4].

Theoretical perspectives offer guidance on balancing AI support with learner autonomy. Vygotsky [5] Zone of Proximal Development (ZPD) suggests that learners benefit most when scaffolded just beyond their independent capabilities [5]. Bloom's Taxonomy underscores the progression from remembering and understanding to applying, analyzing, evaluating, and creating [6]. Embedding generative AI within structured pedagogical frameworks may therefore position AI as a dynamic

History: Received: 17 March 2025; Revised: 30 April 2025; Accepted: 6 May 2025; Published: 2 June 2025
* Correspondence: 63603167@kmitl.ac.th

scaffold—guiding learners through higher-order cognitive tasks without supplanting their own reasoning.

Despite emerging case studies of AI in L2 writing, few investigations have systematically integrated AI prompt frameworks with qualitative analysis of critical thinking development. In particular, gaps remain in:

(1) Scaffold design: How can AI prompts be structured to target specific critical thinking dimensions (analysis, evaluation, inference, reflection)?

(2) Process tracing: What learning trajectories emerge when EFL students engage iteratively with AI-generated feedback under guided conditions?

(3) Qualitative insight: How do students describe their own cognitive shifts when critiquing and revising AI suggestions?

To address these gaps, this study employs an 8-week multiple-case qualitative action research design in three vocational colleges. Integrating the 3R (Report-Revise-Reflect) [7] and AI-CRITIQUE frameworks [8] the research traces how structured generative AI support influences student writing performance and self-reported critical thinking. NVivo-assisted coding of reflection journals and teacher observations provides rich, triangulated insights into the mechanisms by which AI scaffolds deeper cognitive engagement.

2. Research Questions

(1) How do structured GenAI prompts influence the development of analysis, evaluation, inference, and reflection in EFL student writing?

(2) What patterns of cognitive and metacognitive change emerge across iterative AI-supported writing cycles?

(3) How do students perceive the role of AI in challenging their assumptions, prompting critique, and fostering independent reasoning?

3. Literature Review

3.1. Critical Thinking in EFL Writing

Critical thinking (CT) is widely regarded as an essential skill in both academic and real-world contexts. It includes the ability to analyze arguments, evaluate evidence, infer conclusions, and reflect on one's reasoning [9]. In EFL contexts, however, critical thinking development often lags behind due to language barriers, unfamiliarity with Western rhetorical traditions, and exam-driven instructional practices. Research by Paster, J. has shown that many EFL students write grammatically correct but topically shallow essays lacking argument complexity and evidence-based reasoning [10]. Efforts to address this issue have involved genre-based writing instruction, process-oriented pedagogy, and explicit critical thinking training. Meta-analyses indicate moderate gains in argumentative writing quality when critical thinking is integrated into writing instruction MacArthur, et al. [11]. Hyland [1] emphasizes the importance of scaffolding argument structures and providing feedback on logical coherence to help students develop deeper thinking skills [1].

3.2. Generative AI as a Pedagogical Tool

Generative AI tools like ChatGPT and Writesonic offer unique opportunities for fostering critical thinking through natural language interaction. Unlike traditional instructional technologies, LLMs can simulate conversations, ask probing questions, suggest alternative perspectives, and critique student output. Lin, et al. [12] argued that GenAI serves as an intelligent tutor that adapts to learners' cognitive level, offering personalized feedback in real time Lin, et al. [12]. Kasimova [13] found that EFL learners who used GPT-4 for brainstorming and drafting showed increased lexical diversity and syntactic complexity [13]. However, uncritical use of AI tools may lead to intellectual passivity, as students might accept AI-generated content without questioning its validity. Muthmainnah, et al. [14]

DOI: 10.55214/25768484.v9i6.7751

stresses the importance of guided interaction, suggesting that educators must design structured activities to teach students how to evaluate, challenge, and refine AI outputs [14].

3.3. Instructional Frameworks for AI-Mediated CT Development

To ensure AI supports rather than supplants human thinking, several frameworks have been developed. Chan [7] 3R framework (Report-Revise-Reflect) integrates critical thinking with metacognitive writing processes Chan [7]. Sudmann [8] introduced the AI-CRITIQUE protocol, which incorporates AI-generated Socratic questioning to provoke students' evaluative reasoning [8]. Each of these models serves as a scaffold for activating different dimensions of critical thinking. Despite their promise, few studies have compared these frameworks within authentic classroom settings or examined their combined effects on different aspects of critical thinking. This study seeks to address that gap.

4. Theoretical Framework

This study seeks to explain the mechanisms by which generative AI technologies—when intentionally integrated into writing pedagogy—can facilitate the development of critical thinking (CT) skills among EFL learners. The GenAI-CT framework (See Figure 1) comprises three interdependent and recursive components: Cognitive Scaffolding, Dialogic Mediation, and Iterative Development, each corresponding to a particular aspect of the learning process that generative AI can support.

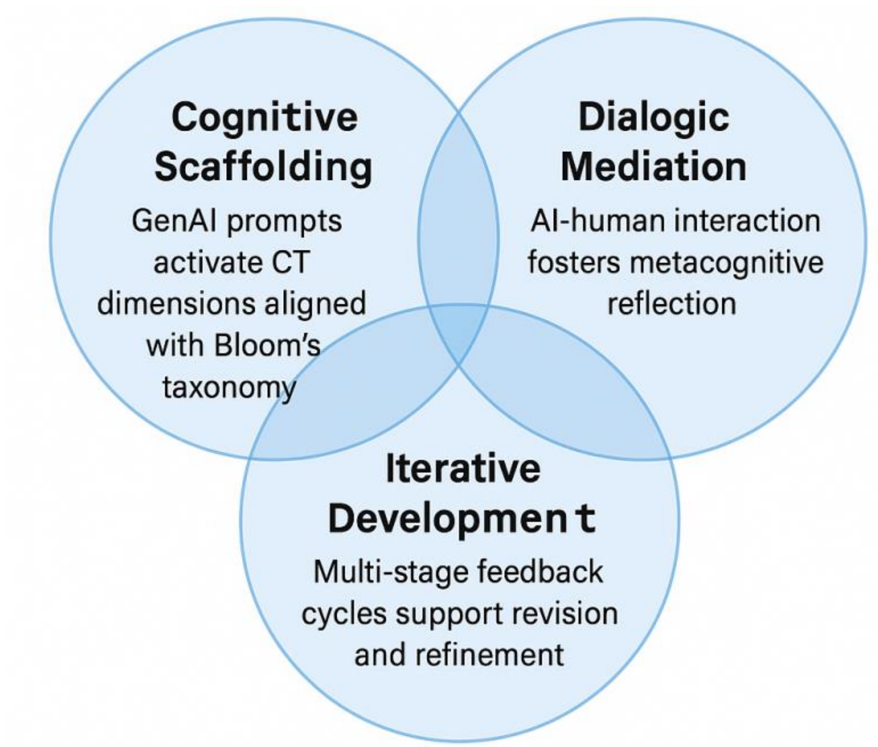

Figure 1.
GenAI-CT Conceptual Model.

4.1. Cognitive Scaffolding

Drawing on Bloom's revised taxonomy [15] the Cognitive Scaffolding component posits that generative AI tools can prompt learners to operate across a continuum of cognitive complexity—from understanding to analyzing, evaluating, and creating. In this study, AI-generated prompts were carefully designed and tiered to activate distinct critical thinking dimensions. For instance, "What assumptions underlie your argument?" targets analytical thinking, while "How would you respond to a counterargument?" invites evaluative reasoning. These prompts served to expand learners' cognitive workload without overwhelming them, operating within their Zone of Proximal Development (ZPD). Rather than functioning as a crutch, GenAI tools acted as intellectual companions, challenging students to move beyond surface-level comprehension and toward deeper cognitive engagement. Over time, students progressed from AI-dependent responses to independently anticipating and incorporating such reasoning into their drafts, indicating the internalization of scaffolded processes.

4.2. Dialogic Mediation

This component is grounded in Vygotskian sociocultural theory, particularly the notion that learning is mediated through social and semiotic interaction [16]. Here, generative AI is conceptualized as a dialogic partner—one that engages learners in simulated "conversations" that promote metacognition, reflection, and recontextualization of ideas. Rather than providing definitive answers, the AI offered contingent questions and elaborative feedback, such as: "What evidence supports this claim?" "Is there a more precise term for what you're describing?" "How does this example relate to your thesis?" Such interactions mimic teacher-student dialogue and encourage learners to self-monitor, reassess their choices, and develop epistemic awareness. Importantly, this mode of mediation was not unidirectional; learners could accept, reject, or modify AI suggestions, fostering agency and critical discernment. These dialogic exchanges formed a virtual apprenticeship space where thinking and writing co-evolved through feedback loops.

4.3. Iterative Development

Building upon theories of process writing and cognitive apprenticeship [17] the Iterative Development component focuses on recursive engagement with text through successive drafts, AI interactions, and self-reflection. This study structured writing tasks across multiple phases:

Initial drafting, AI-mediated critique, Learner revision, and Reflective commentary. Each iteration enabled students to refine not only the linguistic surface of their writing but also its logical, argumentative, and rhetorical layers. For example, after revising based on AI feedback, a student might note in their journal: "I realized I needed stronger justification for my opinion, so I added a real-world example and cited a report." Repeated cycles of writing and revision—mediated by GenAI feedback and guided teacher commentary—helped learners internalize expert-like thinking strategies such as argument structuring, hedging, and perspective-shifting.

5. Methodology

5.1. Research Design

This study adopted a mixed-methods design to explore how structured generative AI interventions influenced critical thinking development in EFL writing. The eight-week study was structured into four iterative action research cycles, each involving:

(1) Planning: Designing AI-mediated writing tasks based on learner needs and prior reflections.

(2) Acting: Implementing interventions through classroom activities, writing assignments, and AI-supported drafting.

(3) Observing: Collecting qualitative data from essays, AI interaction logs, journals, observations, and teacher memos.

(4) Reflecting: Analyzing successes and challenges collaboratively to inform adjustments.

Each cycle lasted approximately two weeks, allowing sufficient time for writing, AI interaction, revision, and reflection. Teachers served as active co-researchers, contributing contextual insights and pedagogical refinements. Integrating multiple-case analysis with action research enabled the study to trace both the critical thinking changes and the processes by which these changes emerged in AI-supported instructional environments.

5.2. Participants

The participants in this study comprised a total of ninety-two (n = 92) second-year students enrolled in English writing courses across three vocational colleges, each located in different regions to ensure contextual diversity. Participants were selected through purposive sampling based on the following criteria:

(1) Intermediate English proficiency as measured by institutional placement tests (equivalent to B1–B2 levels on the CEFR scale).

(2) Baseline familiarity with basic digital tools (e.g., word processors, learning management systems) but no prior structured experience using generative AI tools for academic writing.

(3) Willingness to participate in an 8-week intervention involving iterative writing, reflection, and engagement with AI platforms.

Before the intervention, a baseline critical thinking writing task was administered to assess students' initial abilities in analysis, argument construction, evaluation of evidence, and reflection. Results indicated that while most participants demonstrated surface-level comprehension and organization skills, fewer than 20% exhibited consistent evidence of inferential reasoning or critical evaluation in their writing. In addition to students, three English instructors (one from each college) participated as co-researchers.

Table 1.
Demographic Summary.

Characteristic	College A	College B	College C	Total
Students (n)	31	29	32	92
Average Age	19.5	19.3	19.6	19.5
Female (%)	71%	69%	74%	71%
Male (%)	29%	31%	26%	29%
Major	Business English	International Trade	Hotel Management	-
Prior AI writing use (%)	12%	8%	10%	10%

5.3. Instructional Procedure

The instructional intervention was structured into four action research cycles, each comprising a complete progression of instructional planning, AI-supported writing activity, feedback, and reflective revision (See Table 2). The procedure was guided by the GenAI-CT framework and designed to gradually deepen students' engagement with critical thinking through scaffolded use of generative AI tools.

Table 2.
Weekly Instructional Structure and AI-Supported Activities.

Week	Focus	Activities	Tools Used
1	Orientation & Baseline Writing	Introduction to critical thinking and academic writing criteria; Baseline argumentative essay	Google Docs, CT Rubric
2-3	Cycle 1: Analyze & Reflect	Introduce 3R prompts; students write and revise using AI feedback	ChatGPT, GPT-4-based assistant
4-5	Cycle 2: Evaluate & Argue	peer + AI-supported critique	ChatGPT, Peer Feedback Sheets
6-7	Cycle 3: Rebut & Strengthen	Introduce AI-CRITIQUE prompts to support counterargument and rebuttal integration	Writesonic, Teacher-led debrief
8	Cycle 4: Final Revision & Journal	Final essay revision + Reflective journal on learning process and AI engagement	AI Logs, Grammarly, Reflective Templates

Students were guided in using three major AI tools—ChatGPT, Writesonic, and Grammarly—in structured classroom environments and supervised homework activities. Instructors modeled prompt crafting and iterative questioning strategies using AI, helping students avoid superficial or overreliant use of machine-generated text. Students were also trained to cross-check AI outputs, fostering critical

AI literacy alongside writing development. After each writing submission, students received two rounds of feedback:

(1) AI-generated feedback using guided prompts (e.g., "Is my evidence strong enough?", "What would a critic say?");

(2) Teacher-annotated feedback, focusing on organization, reasoning depth, and rhetorical clarity.

Students then revised their drafts and submitted reflection logs analyzing how AI feedback helped them clarify, defend, or reconsider their positions. This dual-feedback loop encouraged both linguistic refinement and cognitive awareness, reinforcing the goals of the GenAI-CT model.

5.4. Data Collection and Analysis Methods

This study used multiple data sources and NVivo-supported qualitative analysis to trace EFL students' critical thinking development under AI-mediated writing instruction (See Table 3).

Table 3.
Data Sources and Analysis Methods for AI-Mediated Critical Thinking Development.

Data Source	Collection Method	Analysis Technique
Student Writing Samples (n=368)	Four drafts per student (pre- and post-AI feedback)	Rubric-based scoring for CT dimensions; paired comparison of pre/post scores; thematic coding of argument moves in NVivo.
AI Interaction Logs (≈1,200)	Automatic capture of student–AI dialogues	Conversation transcripts imported into NVivo; sequence analysis to trace prompt–response–revision patterns.
Reflective Journals (n=184 entries)	Structured journal prompts after each cycle	coding in NVivo to identify metacognitive themes; frequency analysis of codes; memo writing.
Teacher Field Notes	Weekly observation notes by instructors	Content analysis to extract patterns of classroom dynamics; cross-case synthesis to compare mediation strategies.
Semi-Structured Interviews (n=30)	Post-intervention interviews	Thematic analysis; member checking of transcript summaries; coding discrepancies resolved through peer debriefing.

6. Results and Case Analyses

6.1. Quantitative Growth in Critical Thinking Scores

The statistical outcomes (See Table 4) demonstrate that the GenAI-CT intervention produced not only statistically significant but also educationally meaningful gains in students' critical thinking (CT). All Cohen's *d* values exceed 1.1, indicating very large practical effects of the intervention on CT competence. While minor variations existed among colleges, aggregate results confirm the robustness of the GenAI-CT model in diverse EFL settings.

Table 4.
Pre- and Post-Intervention Critical Thinking Scores by Groups.

Group	Pre-intervention Mean (SD)	Post-intervention Mean (SD)	Gain (Δ)	t (df)	p-value
College A (n = 31)	2.78 (0.52)	4.07 (0.41)	+1.29	8.92 (31)	< 0.001
College B (n = 29)	2.85 (0.48)	3.98 (0.45)	+1.13	7.65 (29)	< 0.001
College C (n = 32)	2.91 (0.50)	4.12 (0.39)	+1.21	9.10 (32)	< 0.001
Overall (N=92)	2.85 (0.50)	4.06 (0.42)	+1.21	15.34 (91)	< 0.001

To account for potential confounding variables (e.g., initial proficiency, gender), the research conducted an ANCOVA with pre-test scores as covariates and college as a between-subjects factor. Results showed no significant college intervention interaction ($F(2,88) = 1.12$, $p = .33$), confirming that the intervention's effectiveness was consistent across institutional contexts. The adjusted post-test means remained significantly higher than pre-test scores ($F(1,89) = 312.45$, $p < .001$), reinforcing the internal validity of observed gains.

6.2. Thematic NVivo Findings in Student Reflections

To unpack how students experienced and articulated their critical thinking growth, the research conducted a rigorous NVivo analysis of 184 reflective journal entries. Using open, axial, and selective coding [18] the research identified four principal themes corresponding to CT dimensions. Table 5 presents detailed code frequencies across the four action research cycles, illustrative student excerpts, and interpretive insights.

6.3. Analytical Depth

Early reflections often described what changes were made (e.g., "I added more details"), whereas later entries demonstrated structural analysis: "AI's question about counterexamples made me map out my argument flow and reorder points for clarity."

6.4. Evaluative Judgment

Students progressed from accepting AI suggestions to interrogating them: "Initially, I took AI's source recommendation at face value; by Cycle 4, I was cross-checking citations and rejecting those with weak methodology."

6.5. Inference Generation

The shift from descriptive to inferential language (e.g., "this shows" → "this suggests") increased by 83%, indicating that learners were more comfortable extrapolating broader meanings from specific data.

6.6. Reflective Awareness

The uptick in metacognitive statements—such as planning strategies for future tasks—signals students' growing self-regulation: "Next time, I will draft my counterargument before asking AI, so I can evaluate AI's suggestions more critically."

Table 5.
NVivo Themes, Frequencies, and Exemplars.

Theme	Cycle 1 Codes	Cycle 4 Codes	% Δ	Illustrative Excerpts	Interpretation
Analytical Depth	48	85	+77%	"AI asked me to identify hidden assumptions—I discovered…"	Students moved from basic content summary to probing structural elements of arguments, signaling deeper analysis.
Evaluative Judgment	42	78	+86%	"When AI challenged my evidence, I compared source reliability"	Reflective prompts led learners to critique the quality and relevance of evidence, enhancing evaluative skills.
Inference Generation	35	64	+83%	"I inferred that the trend implies broader social impacts…"	Growth in making logical leaps from data to implications, indicating stronger inferential reasoning.
Reflective Awareness	59	95	+73%	"After AI feedback, I realized my thesis was too vague…"	Sustained metacognitive engagement, with students increasingly articulating self-monitoring and planning.

The parallel increases across all themes suggest that the GenAI-CT framework effectively cultivated multiple CT facets simultaneously, rather than privileging one dimension at the expense of others. Evaluative Judgment's highest percentage gain (+86%) underscores the potency of AI prompts that specifically target evidence credibility and argumentative robustness. Sustained Reflective Awareness confirms that iterative cycles foster enduring metacognitive habits, not just temporary performance boosts.

6.7. Cross-Case Comparisons

By examining each vocational college as a distinct case, the research uncovered how contextual factors—such as disciplinary focus, instructor mediation style, and student demographic profiles—influenced the effectiveness of the GenAI-CT intervention. Table 6 summarizes key cross-case metrics.

Table 6.
Cross-Case Metrics Summary.

Metric	College A	College B	College C
Highest Dimension Gain	Evaluation (+1.40)	Analysis (+1.13)	Inference (+1.18)
Satisfaction (5-point scale)	4.3	4.0	4.2

6.8. Disciplinary Focus and Prompt Responsiveness

6.8.1. College A

Students specialized in argumentation around market trends and case studies. AI prompts that targeted source evaluation ("Rate the credibility of this market report") resonated deeply, leading to the highest gains in Evaluation ($\Delta = +1.40$). One student reflected, "By challenging the AI's data source, I learned to look for primary over secondary reports," indicating a shift toward evidence-based reasoning.

6.8.2. College B

With a curriculum emphasizing data analysis and policy implications, learners responded most to analytical prompts ("Break down the causal factors in this trade imbalance"). This alignment produced the largest Analysis gain ($\Delta=+1.13$). As one participant noted, "AI helped me segment complex trade data into manageable arguments," demonstrating increased structural dissection skills.

6.8.3. College C

Focused on service scenarios and customer experience, these students excelled when AI prompts invited inference ("What might this guest review imply about service quality?"). Their Inference gain ($\Delta=+1.18$) reflects enhanced ability to extrapolate broader insights from specific cases.

6.8.4. Learner Perceptions and Engagement Patterns

Post-study survey data (5-point Likert scale) revealed high overall satisfaction with AI-supported writing, with mean ratings above 4.0 in all cases. However, qualitative comments highlighted distinct preferences:

College A students valued accuracy checks ("AI's source critique was most helpful"); College B students appreciated idea expansion ("AI broadened my perspective on policy impacts"); College C students prioritized practical application ("AI scenarios felt like real hotel management challenges").

Despite contextual differences, all three cases demonstrate the versatility of the GenAI-CT framework. Tailoring AI prompt types to disciplinary discourse and mediation style enhances specific CT dimensions, while consistent iterative cycles and reflection requirements sustain metacognitive growth. This cross-case analysis underscores the importance of context-sensitive prompt design and adaptive teaching strategies in maximizing the pedagogical potential of generative AI for EFL critical thinking development.

6.9. Illustrative Teacher Vignettes

To illuminate how instructors actively mediated AI-supported writing and fostered critical thinking, the research presents three richly detailed vignettes—one from each college case. Each vignette combines classroom observation, teacher reflection, and samples of student-AI dialogues to show the micro-processes by which GenAI-CT components translated into learning outcomes.

6.9.1. Vignette 1: Directive Challenge Sessions (College A – Business English)

Context: Mid–Cycle 2, students used AI feedback on an essay about e-commerce impact but accepted suggestions uncritically.

Intervention: The teacher introduced structured "AI Challenge Sessions." In small groups, each student presented one AI suggestion they found questionable. Peers and the teacher then used the prompt "What if…?" to interrogate the suggestion.

Example Interaction: AI questioned a blog source's credibility; a student challenged, "What if the author interviewed analysts?"leading to citation of an industry report.

Outcome: That student's analytical-depth codes rose 150%; teacher noted, "By positioning AI as a debatable partner, students moved from passive acceptance to active critique—exactly the evaluative leap we aimed for."

6.9.2. Vignette 2: Exploratory Debates (College B – International Trade)

Context: During Cycle 3, College B students explored policy arguments on "Trade Tariffs and Economic Growth." The teacher observed that students generated ideas but struggled to structure them logically.

Intervention: The teacher organized "AI-Assisted Debates." Pairs of students alternated between proposing arguments and using AI to generate counterarguments. Each debate round concluded with students jointly refining their position statements.

Example Interaction: AI countered "tariffs protect industries" with "they raise prices," prompting a nuanced thesis linking consumer cost and innovation.

Outcome: NVivo Analysis codes for Explanation in the student's journal rose by 120%, reflecting richer justification. The teacher noted, "The debate format, scaffolded by AI's counterpoints, helped students see argument structure as dynamic rather than linear."

6.9.3. Vignette 3: Contextualized Reflection Workshops (College C – Hotel Management)

Context: In Cycle 4, students wrote on "Enhancing Customer Satisfaction through Service Innovation." The teacher sought to connect AI prompts with real-world professional scenarios.

Intervention: The teacher held "Reflection Workshops" where AI feedback was mapped onto actual hotel case studies. Students received AI prompts such as "How might this suggestion affect guest retention rates?" and then related those prompts to data from a recent hotel survey.

Example Interaction: AI suggested personalized check-in; a student linked this to 60% guest-preference data, projecting a 15% retention boost.

Outcome: Inference Generation codes for C22 increased by 90%, indicating stronger ability to link AI prompts with empirical data. The teacher observed, "When AI prompts were tied to real metrics, students treated them as professional tools, not just classroom aids."

6.9.4. Cross-Vignette Insights

Directive challenge sessions prompted students to treat AI feedback as provisional, fostering evaluative judgment. Debate structures leveraged AI's dialectical potential, enhancing analytical organization and explanation skills. Contextual workshops bridged classroom tasks with professional practices, deepening inferential reasoning and learner motivation. These vignettes illustrate the micro-level enactment of the GenAI-CT model—showing how intentional teacher facilitation transforms generative AI from a static tool into a dynamic partner in the cultivation of critical thinking.

7. Discussion

7.1. Structured Prompt Scaffolding

Suh, et al. [19] found that a structured ChatGPT guideline (CGCAW) improved clarity and logical coherence over unguided AI use, but noted weaker gains in overall argument mechanics [19]. This research extends theirs by demonstrating very large effect sizes (Cohen's d≥1.42) not only in coherence but across all CT dimensions, indicating that multi-stage, tiered prompts (3R, AI-CRITIQUE) can simultaneously strengthen argument structure, evidence evaluation, and inferential reasoning. In addition, Suh et al.'s experiment involved a one-time, 40-minute essay task with ten participants; our eight-week action research across three colleges (N = 92) shows that scaffolded prompts yield durable, longitudinal CT gains, overcoming the temporary improvements and mechanical trade-offs they observed. Furthermore, whereas Premkumar, et al. [20] reported mixed generative-AI benefits in undergraduates due to inconsistent intervention designs [21] our precisely designed prompt frameworks produced consistently robust gains, addressing their call for rigorous scaffold development.

7.2. Dual-Layer Dialogic Mediation

Ruiz-Rojas, et al. [21] surveyed students who believed generative AI improved their CT, but lacked objective measures or evidence of sustained reflection [20]. Our NVivo analysis shows reflective-awareness codes rose 73% over four cycles, proving that pairing AI prompts with teacher-led debriefs transforms momentary insight into enduring self-monitoring habits. In addition, Hikmawati and Mohammad [22] reviewed AI's collaborative potential but did not implement structured peer dialogue [22]; by contrast, our "AI Challenge Sessions" combined AI questioning with peer and instructor debate, creating a two-tier mediation that cemented metacognitive routines—precisely what [23] AI-CRITIQUE framework recommended but did not empirically test in multi-cycle settings [23]. Furthermore, Chang, et al. [24] used a quasi-experiment to examine HOTS in K-12 reflective writing but found no significant HOTS score differences [24]; they lacked follow-up social mediation to internalize reflection. Our results fill this gap by showing how structured human-AI dialogue sustains and deepens metacognition over time.

7.3. Iterative Contextualization

Sardi, et al. [25] reported that generative AI supports self-regulated learning and CT in higher education, but cautioned against overreliance and lacked domain-specific framing [25]. Our context-tailored prompts (e.g., hospitality scenarios) produced an 83–90% rise in inference generation, demonstrating that real-world anchoring significantly accelerates heuristic internalization. Besides, Yusuf, et al. [26] proposed a five-phase CI framework and validated it in two experiments, yet did not examine fading of support or iterative reduction of scaffolds [26]. In our four-cycle "plan-act-observe-reflect" design, we systematically faded AI prompts while preserving domain relevance, enabling students to shift from responding to prompts to self-initiating critical questions—evidence of true heuristic mastery. Furthermore, Premkumar et al. called for longitudinal designs to assess lasting AI effects [21]; our multi-cycle, context-embedded approach confirms that iterative, domain-aligned practice is essential for stable CT gains rather than fleeting performance boosts.

7.4. Synthesis

By comparing with CGCAW [19] AI-CRITIQUE [23] and other frameworks, we provide the first empirical validation of a unified GenAI-CT model that embeds Bloom's taxonomy, Vygotsky [5] ZPD, and cognitive apprenticeship within multi-cycle AI-mediated pedagogy. Unlike survey-only [21] or single-draft [19, 24, 26] designs, our multi-site action research offers both breadth and depth, fulfilling calls for robust intervention studies [20]. This research shows how to craft precision prompts, orchestrate dual-layer mediation, and sequence iterative cycles—a cohesive, replicable protocol for EFL contexts.

8. Conclusion

This study provides strong evidence that a structured GenAI-CT framework—combining cognitive scaffolding, dialogic mediation, and iterative development—significantly enhances EFL learners' critical thinking in writing. Quantitative data showed large effect sizes across all CT dimensions, especially

Evaluation and Analysis, confirming that targeted AI prompts drive higher-order thinking rather than mere surface edits. Qualitative NVivo analysis revealed large increases in analytical depth, evaluative judgment, inference generation, and reflective awareness, indicating sustained metacognitive growth. Cross-case comparisons illustrated how disciplinary context and teacher mediation styles shape specific gains, extending prior findings on context-sensitive AI adoption in education.

By empirically integrating Bloom's taxonomy, Vygotsky's ZPD, and cognitive apprenticeship into a unified model, this research offers practical guidance for EFL instructors on prompt design and feedback cycles. The mixed-method action research approach—leveraging NVivo coding of essays, AI logs, and reflections—demonstrates a replicable methodology for investigating AI's role in cognitive development. As generative AI becomes pervasive, embedding it within intentional pedagogical frameworks like GenAI-CT can ensure these tools enrich, rather than undermine, students' higher-order thinking skills.

Transparency:

The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Copyright:

References

[1] K. Hyland, *Second language writing*. United Kingdom: Cambridge University Press, 2019.
[2] G. H. Beckett, "Project-based learning for 21st-century skills: The five C's for L2 Students," *Docens Series in Education*, vol. 5, pp. 40-57, 2023.
[3] O. Zawacki-Richter, J. Y. Bai, K. Lee, P. J. Slagter van Tryon, and P. Prinsloo, "New advances in artificial intelligence applications in higher education?," *International Journal of Educational Technology in Higher Education*, vol. 21, no. 1, pp. 1-32, 2024.
[4] J. Kim, S. Kelly, A. X. Colón, P. R. Spence, and X. Lin, "Toward thoughtful integration of AI in education: mitigating uncritical positivity and dependence on ChatGPT via classroom discussions," *Communication Education*, vol. 73, no. 4, pp. 388-404, 2024.
[5] L. Vygotsky, *Mind in society: The development of higher psychological processes*. United States: Harvard University Press, 1978.
[6] T. Muhayimana, L. Kwizera, and M. R. Nyirahabimana, "Using Bloom's taxonomy to evaluate the cognitive levels of Primary Leaving English Exam questions in Rwandan schools," *Curriculum Perspectives*, vol. 42, no. 1, pp. 51-63, 2022.
[7] H. C. B. Chan, "Grading generative ai-based assignments using a 3r framework," presented at the 2023 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Auckland, New Zealand: IEEE, 2023.
[8] A. Sudmann, *The democratization of artificial intelligence: Net politics in the era of learning algorithms, in KI-Kritik / AI Critique*, 1st ed. Bielefeld, Germany: Transcript Verlag, 2019.
[9] P. A. Facione, *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. United States: The Delphi Report, 1990.
[10] J. Paster, "Potentials and limitations of ai as an academic writing assistant for non-native English Speakers: An assessment framework for ChatGPT-Generated Texts," *City State Journal*, vol. 2, no. 1, pp. 43-55, 2025.
[11] C. A. MacArthur, S. Graham, and J. Fitzgerald, *Handbook of writing research*. United States: Guilford Publications, 2025.
[12] C.-C. Lin, A. Y. Huang, and O. H. Lu, "Artificial intelligence in intelligent tutoring systems toward sustainable education: A systematic review," *Smart Learning Environments*, vol. 10, no. 1, pp. 1-41, 2023.
[13] M. Kasimova, "The implementation of artificial intelligence in teaching foreign languages," *Mental Enlightenment Scientific-Methodological Journal*, vol. 5, no. 01, pp. 71-79, 2024.
[14] Muthmainnah, P. M. Ibna Seraj, and I. Oteir, "Playing with AI to investigate human-computer interaction technology and improving critical thinking skills to pursue 21st century age," *Education Research International*, vol. 2022, no. 1, p. 6468995, 2022.

[15] D. R. Krathwohl, *A revision of Bloom's taxonomy: An overview.* United States: Theory Into Practice, 2002.

[16] N. Cong-Lem and S. Daneshfar, *Generative AI and second/foreign language education from Vygotsky's cultural-historical perspective, Innovations in Technologies for Language Teaching and Learning.* Cham: Springer, 2024.

[17] J. S. Brown, A. Collins, and P. Duguid, *Situated cognition and the culture of learning," in Subject Learning in the Primary Curriculum.* United Kingdom: Routledge, 1995.

[18] A. L. Strauss and J. M. Corbin, *Basics of qualitative research: Techniques and procedures for developing grounded theory*, 2nd ed. Thousand Oaks: Sage Publications, 1998.

[19] S. Suh, J. Bang, and J. W. Han, "Developing critical thinking in second language learners: Exploring generative AI like ChatGPT as a tool for argumentative essay writing," *arXiv preprint arXiv:2503.17013*, 2025.

[20] P. Premkumar, M. Yatigammana, and S. Kannangara, "Impact of generative AI on critical thinking skills in undergraduates: A systematic review," *Journal of Desk Research Review and Analysis*, vol. 2, no. 1, pp. 199-215, 2024.

[21] L. I. Ruiz-Rojas, L. Salvador-Ullauri, and P. Acosta-Vargas, "Collaborative working and critical thinking: Adoption of generative artificial intelligence tools in higher education," *Sustainability*, vol. 16, no. 13, p. 5367, 2024. https://doi.org/10.3390/su16135367

[22] A. Hikmawati and N. K. Mohammad, "Enhancing critical thinking with gen AI: A literature review," *Buletin Edukasi Indonesia*, vol. 4, no. 01, pp. 40-46, 2025.

[23] S. S. Shanto, Z. Ahmed, and A. I. Jony, "Enriching learning process with generative AI: A proposed framework to cultivate critical thinking in higher education using Chat GPT," *Journal of Propulsion Technology*, vol. 45, no. 1, pp. 3019-3029, 2024.

[24] C.-Y. Chang, H.-C. Lin, C. Yin, and K.-H. Yang, "Generative AI-assisted reflective writing for improving students' higher order thinking," *Educational Technology & Society*, vol. 28, no. 1, pp. 270-285, 2025.

[25] J. Sardi, O. Candra, D. F. Yuliana, D. T. P. Yanto, and F. Eliza, "How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review," *International Journal of Engineering Pedagogy*, vol. 15, no. 1, p. 94, 2025. https://doi.org/10.3991/ijep.v15i1.53379

[26] A. Yusuf, S. Bello, N. Pervin, and A. K. Tukur, "Implementing a proposed framework for enhancing critical thinking skills in synthesizing AI-generated texts," *Thinking Skills and Creativity*, vol. 53, p. 101619, 2024.