BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies
https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Cognitive Offload Instruction with Generative AI: A Quasi-Experimental Study on Critical Thinking Gains in English Writing

Hui Hong*, ***Poonsri Vate-U-Lan***, ***Chantana Viriyavejakul***

School of Industrial Education and Technology , King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

ABSTRACT

This study explores the impact of generative AI-enabled cognitive offload instruction on the development of critical thinking skills in English essay writing among first-year university students. A quasi-experimental design was employed, comparing traditional instruction with an AI-augmented pedagogy that delegated lower-order writing tasks to generative AI tools, allowing students to focus on analysis, evaluation, and reflection. Over 12 weeks, 240 participants engaged in structured writing cycles involving AI brainstorming, individual critique, peer-AI co-revision, and reflective journaling. Results revealed that the AI-enabled cognitive offload group demonstrated significantly greater improvements in standardized critical thinking assessments and produced higher-quality essays in terms of logical coherence, evidence use, and originality. Mediation analysis indicated that cognitive offloading behavior partially explained the relationship between AI use and critical thinking gains. The findings suggest that when generative AI is integrated into pedagogy through deliberate scaffolding, it can enhance rather than hinder higher-order thinking. This study highlights the importance of balancing technological efficiency with instructional strategies that promote active engagement, metacognitive reflection, and collaborative learning. It offers practical implications for educators seeking to incorporate AI tools without compromising the development of essential cognitive skills, proposing that structured cognitive offload instruction can serve as an effective approach to fostering critical thinking in second-language writing contexts.

Keywords: Critical Thinking; Cognitive Offloading; Generative AI; English Essay Writing

*CORRESPONDING AUTHOR:
Hui Hong, School of Industrial Education and Technology , King Mongkut's Institute of Technology Ladkrabang; Email: 63603167@kmitl.ac.th

ARTICLE INFO
Received:18 May 2025 | Revised:9 June 2025 | Accepted: 20 June 2025 | Published Online: 15 July 2025
DOI: https://doi.org/10.30564/fls.v7i7.10072

CITATION
Hong, H., Lastname, Vate-U-Lan, P., Viriyavejakul, C., 2025. Cognitive Offload Instruction with Generative AI: A Quasi-Experimental Study on Critical Thinking Gains in English Writing. Forum for Linguistic Studies. 7(7): 325–334. DOI: https://doi.org/10.30564/fls.v7i7.10072

1. Introduction

This study situates itself at the intersection of second-language writing pedagogy, cognitive psychology, and educational technology, aiming to address a pressing need for instructional designs that leverage generative AI without undermining learners' higher-order thinking. Over the past decade, L2 writing instruction has evolved from process-oriented approaches toward more metacognitively rich frameworks that foreground students' analytical and evaluative capacities. Meanwhile, generative AI tools such as ChatGPT have demonstrated remarkable abilities to automate lower-order writing tasks, offering new affordances for cognitive offloading.

Cognitive offloading, which involves delegating routine tasks to external resources like AI tools, can free up mental resources for more complex cognitive processes [1]. Research has shown that AI-supported scaffolding systems help students generate stronger arguments by enabling a focus on higher-order thinking. For instance, Kim et al. demonstrated that such AI scaffolds improved students' argumentative depth [2], while Gerlich highlighted the risks of unstructured AI use, which can lead to "cognitive laziness" in the absence of adequate support [3]. Critical Thinking (CT) is a vital skill in L2 writing, involving interpretation, analysis, evaluation, and self-regulation. Studies such as Huang and Zhang have shown that process-genre instruction can significantly improve critical thinking in writing [4]. Moreover, prior research on AI in writing instruction emphasizes the benefits of structured AI use. Yang et al. demonstrated that scaffolded AI interventions significantly improved essay quality [5], while Yin and Jiahao proposed a hybrid AI framework that enhanced critical thinking and prompt-engineering skills [6]. These studies highlight the importance of scaffolding when leveraging AI tools to develop critical thinking.

Despite this promise, a critical gap remains: unstructured use of AI can foster "cognitive laziness," diminishing students' engagement with analysis, synthesis, and reflection. Prior investigations have documented negative correlations between AI reliance and critical thinking outcomes when no pedagogical scaffolds were provided [7]. What remains underexplored is how intentionally structured "cognitive offload instruction", in which generative AI handles routine tasks while learners focus on higher-order processes, might mitigate these risks and, enhance critical thinking.

Accordingly, this study seeks to fill three interrelated gaps. First, although cognitive offloading theory suggests that delegating routine work can free mental resources for complex thinking, empirical evidence is lacking in authentic L2 writing contexts. Second, while educators have proposed active learning strategies for AI integration, systematic comparisons between AI-augmented offload instruction and traditional pedagogy remain scarce. Third, the mediating role of offloading behavior in the AI and critical thinking relationship has not been quantitatively tested in essay writing settings. To address these gaps, the present study poses the following research questions:

1. What effect does generative AI—enabled cognitive offload instruction have on students' critical thinking gains in English essay writing compared to traditional instruction?
2. How does essay quality—measured in logical coherence, evidence use, and originality—differ between AI-augmented and control groups?
3. To what extent does cognitive-offloading behavior mediate the relationship between AI-enabled instruction and critical thinking outcomes?

2. Literature Review

2.1. Critical Thinking in L2 Writing

Critical Thinking (CT) in second-language (L2) writing has been recognized as a pivotal competence that enables learners to move beyond mere linguistic accuracy toward deeper analysis, synthesis, and evaluation of ideas in their essays. At its core, CT involves skills such as interpretation, analysis, inference, evaluation, explanation, and self-regulation, as articulated by Facione's widely-adopted Delphi definition [8]. In L2 contexts, these skills are complicated by additional demands on language proficiency, making the cultivation of CT both more challenging and more crucial for academic success. Moreover, writing itself has been framed as a form of thinking, underscoring that writing in a foreign language can serve as a catalyst for critical reflection and idea development [9].

Empirical studies have documented both the challenges L2 writers face and the potential for pedagogical intervention to foster critical thinking (CT). A quasi-experimental study in China demonstrated that process-genre instruction significantly enhanced learners' self-reported use of metacognitive strategies and CT in argumentative essays [4]. Additionally, a study in Language Testing in Asia identified key linguistic and rhetorical features—such as logical connectors and evidence elaboration—that reliably signaled CT in L2 argumentative essays, providing empirical grounding for rubric development [10].

While peer review and computer-mediated feedback have been shown to enhance both local and global revision skills and to foster CT through explicit scaffolding (e.g., reviewing, restructuring, explaining), many instructors still lack systematic approaches to integrate CT instruction into L2 writing curricula. Esmaeil Nejad et al. argued that communicative language teaching must be augmented with tasks requiring learners to generate and critically support ideas, yet concrete models for doing so remain scarce [11]. Recent research has called for more fine-grained investigation of how CT development varies by proficiency level, genre, and instructional format.

2.2. Generative AI and Cognitive Offloading

Generative AI tools, such as OpenAI's ChatGPT, Google's Bard, and Anthropic's Claude, have rapidly advanced in their ability to produce coherent text, brainstorm ideas, and perform routine writing tasks on demand. This has led researchers to examine how these systems function as "cognitive prosthetics", enabling users to delegate lower-order mental operations [3] (e.g., idea generation, lexical retrieval, basic drafting) to external algorithms.

Cognitive offloading refers to the process by which individuals reduce their internal cognitive load by relying on external resources, such as notes, calculators, or AI systems, to perform tasks that would otherwise consume working memory capacity. According to the extended mind thesis, these external tools become integrated into one's cognitive system, effectively expanding mental capabilities. When generative AI is used for brainstorming or initial drafting, writers can free up attentional resources for higher-order processes like analysis, evaluation, and reflection [12].

Recent experimental studies have begun to quantify the effects of AI-enabled cognitive offloading on writing performance and cognitive engagement. For instance, Fan et al. conducted a randomized study comparing learners supported by ChatGPT, human experts, writing analytics tools, and no extra tools [13]. The results revealed that while the ChatGPT group outperformed others in essay score improvement, their knowledge gain and transfer were not significantly different from those of other groups, suggesting potential over-reliance on AI tools. To mitigate these risks, instructional designs have incorporated scaffolding strategies, such as requiring users to annotate, critique, and revise AI outputs. These findings underscore the importance of structured AI integration to prevent cognitive atrophy and promote active engagement in the writing process.

Despite growing interest, key gaps remain: most studies focus on single-session tasks in laboratory settings; longitudinal effects in authentic classroom contexts are underexplored; few studies have quantitatively examined how cognitive offloading mediates the relationship between AI use and higher-order outcomes; comparative research on different generative AI platforms and prompt engineering techniques is limited.

2.3. Instructional Strategies with AI

Effective scaffolded prompt design begins by having students generate initial text with AI and then immediately engage in critique and reflection on that output. For example, learners might ask ChatGPT to draft an argumentative paragraph and then respond to metacognitive prompts such as "What assumptions underlie this paragraph?" and "How could the reasoning be improved?" Research shows that embedding such layered questioning and reflection requirements significantly increases students' analytic and evaluative behaviors, leading to higher critical-thinking scores on subsequent assessments [14].

Collaborative AI review leverages both peer interaction and AI assistance in a unified process. In practice, one student operates the AI tool to produce a draft while peers simultaneously review the AI's suggestions, discussing strengths and weaknesses in real time. Studies indicate that this peer-AI co-revision model fosters deeper argumentative depth and collective reflection, as students must articulate evaluative judgments both about the AI's output and their classmates' feedback [15].

Process-oriented workshops integrate AI into a multi-stage writing cycle in which students rotate through distinct phases: AI-driven brainstorming, individual drafting, and group critique, ensuring that AI remains one component of a broader pedagogical sequence rather than a standalone solution. Technology fluency training emphasizes teaching students to craft precise, goal-oriented prompts and to compare outputs across different AI platforms. Instruction in prompt engineering, such as specifying audience, tone, and complexity, requires learners to clarify their own rhetorical intentions, which in turn enhances the sophistication of AI-generated drafts [16]. Additionally, assignments that ask students to evaluate and contrast outputs from tools like ChatGPT and Bard cultivate evaluative judgment and a nuanced understanding of each platform's affordances and limitations [17].

3. Methodology

3.1. Research Design

A nonequivalent groups quasi-experimental design was chosen because full randomization was impractical in intact classroom settings. A pretest–posttest structure with a control group allowed adjustment for baseline differences, reducing bias in effect estimates. Mixed-methods integration occurred at the design level (convergent-parallel), whereby quantitative outcomes were interpreted alongside qualitative patterns of student engagement and offloading behaviors.

3.2. Participants and Sampling

Participants were 240 first-year English majors from four colleges (60 per institution). Eligibility required intermediate-high English proficiency (CEFR B2) and no prior structured AI-writing instruction. Colleges were selected via convenience sampling; within each, two classes were assigned to the AI intervention and the two to control, yielding 120 students per condition. Sample size targets were informed by educational quasi-experiment standards ($n \approx 100$–300) to detect medium effect sizes ($d \approx 0.50$) with 80% power at $\alpha = 0.05$.

3.3. Intervention Procedures

Over 12 weeks, the intervention group participated in AI-enabled cognitive offload instruction. Each weekly session followed a cycle: (1) AI brainstorming (students prompt ChatGPT to generate outlines) [18]; (2) individual analysis (students critique AI suggestions via metacognitive reflection prompts); (3) group revision (peer-AI co-revision workshops) [19]; (4) reflective journaling on offloading decisions [20]. The control group followed a parallel cycle without AI, using traditional brainstorming and peer review. Fidelity was monitored via session logs and instructor checklists.

3.4. Instruments and Measures

(1) Critical Thinking Test: A validated L2-appropriate version of the Watson-Glaser Critical Thinking Appraisal (pre- and post-test) measured inference, recognition of assumptions, deduction, interpretation, and evaluation ($\alpha = 0.89$).

(2) Essay-Quality Rubric: Adapted from the Paul-Elder elements and empirical L2 writing features (logical connectors, evidence elaboration), three blind raters scored essays on coherence, evidence use, and originality (inter-rater $\alpha = 0.92$).

(3) Cognitive Offloading Scale: A six-item, Likert-type questionnaire assessing reliance on AI for specific tasks (e.g., idea generation, lexical choice), adapted from cognitive load measurement literature ($\alpha = 0.90$).

(4) Qualitative Logs & Interviews: Reflection journals and semi-structured interviews probed decision rationales and perceived engagement; transcripts were coded in NVivo 12 for themes of offloading, engagement, and metacognition.

3.5. Data Collection Procedures

Quantitative instruments were administered online under proctored conditions at weeks 1 and 12. Essays were collected the via LMS and anonymized. Offloading scales were delivered immediately after each writing task to capture situational reliance. At mid-intervention and post-intervention, a purposive subsample of 24 students (balanced by group and proficiency) participated in one-on-one interviews, each lasting approximately 30 minutes.

3.6. Data Analysis

Quantitative: ANCOVA compared post-test critical thinking scores while controlling for pre-test scores. Essay-rubric differences were tested via independent-samples t-tests. Mediation analysis (bias-corrected bootstrap, 5,000 samples) examined whether offloading scores mediated the instruction-CT relationship. Assumptions (normality, homogeneity, linearity) were checked via residual plots and Levene's test. Analyses used SPSS 27 and Hayes' PROCESS macro.

Qualitative: NVivo-facilitated thematic analysis followed Braun and Clarke's six-phase approach: familiarization, coding, theme development, review, definition, and reporting. Triangulation across journals, interviews, and observation logs enhanced credibility.

3.7. Reliability, Validity, and Trustworthiness

Instrument reliability was confirmed via Cronbach's alpha (>0.85 for all scales) and inter-rater agreement. Construct validity of the offloading scale was supported by confirmatory factor analysis (CFI = 0.95, RMSEA = 0.04). Qualitative trustworthiness was established through member checking, audit trails, and peer debriefing. Methodological triangulation bolstered internal validity, while detailed context descriptions support transferability.

4. Research Results

4.1. Critical Thinking Gains

In this study, students in the AI-enabled cognitive offload condition demonstrated markedly greater improvements in critical thinking scores than those in the control condition, with a mean gain of 12.3 points versus 0.5 points, respectively. This difference was statistically significant and corresponds to a large effect (partial $\eta^2 = 0.15$), indicating that the structured AI offload intervention explained 15% of the variance in post-test scores.

As shown in **Table 1**, the AI-enabled offload group's mean critical thinking score increased from 66.3 (SD = 9.1) on the pre-test to 78.6 (SD = 8.4) on the post-test, whereas the control group improved marginally from 65.8 (SD = 9.4) to 66.3 (SD = 9.1). An ANCOVA controlling for pre-test scores yielded $F(1, 237) = 42.7$, $p < 0.001$, demonstrating that the intervention effect was highly significant.

According to Cohen's benchmarks for partial η^2, our observed value of .15 qualifies as a large effect, underscoring the practical importance of the AI offload design Cross Validated. This magnitude exceeds typical medium-sized gains ($d \approx 0.50$) reported in shorter-term scaffolded AI writing interventions.

Table 1. Critical Thinking Scores and ANCOVA Results.

Group	Pre-Test Mean (SD)	Post-Test Mean (SD)	Δ Mean	F (ANCOVA)	*p*-Value	Partial η²
AI-enabled Offload (n = 120)	66.3 (9.1)	78.6 (8.4)	+12.3	F(1, 237) = 42.7	<0.001	0.15
Control (n = 120)	65.8 (9.4)	66.3 (9.1)	+0.5			

Notes:

1. Δ Mean = Post-test Mean—Pre-test Mean.
2. ANCOVA controlled for baseline differences via pre-test covariate.
3. Partial η^2 benchmarks per Cohen: 0.01 = small, 0.06 = medium, 0.14 = large

4.2. Essay Quality

Across six dimensions of essay quality, the AI-enabled cognitive offload group significantly outperformed the control group, with substantial gains in logical coherence, evidence use, originality, lexical sophistication, syntactic complexity, and text cohesion. These results align with prior research showing that scaffolded AI-writing interventions enhance both organizational clarity and argumentative depth compared to unscaffolded AI use or traditional instruction. Observed effect sizes (Cohen's d between 1.1 and 1.4) exceed those reported in short-term AI workshop studies, underscoring the impact of a semester-long scaffolded design in authentic L2 contexts.

As shown in **Table 2**, essays from the AI-offload group scored higher in logical coherence (M = 4.2, SD = 0.5) than those from the control group (M = 3.5, SD = 0.6), $t(238) = 10.1$, $p < 0.001$, indicating more systematic argument flow and clearer transitions. Evidence use, defined as the integration and citation of supporting details, was also stronger in the experimental group (M = 4.0, SD = 0.6) versus the control group (M = 3.2, SD = 0.7), t(238) = 9.3, $p < 0.001$, reflecting more robust engagement with source material. Ratings of originality were higher in the AI-offload group (M = 3.9, SD = 0.6) compared with the control group (M = 3.1, SD = 0.8), t(238) = 8.7, $p < 0.001$, suggesting that students generated more novel arguments when routine tasks were delegated to AI.

Table 2. Essay Quality Metrics and Statistical Comparisons.

Metric	Group	Mean (SD)	t (df = 238)	*p*-Value	Cohen's d
Logical Coherence	AI-Offload (n = 120)	4.2 (0.5)	10.1	<0.001	1.31
	Control (n = 120)	3.5 (0.6)	—	—	—
Evidence Use	AI-Offload	4.0 (0.6)	9.3	<0.001	1.20
	Control	3.2 (0.7)	—	—	—
Originality	AI-Offload	3.9 (0.6)	8.7	<0.001	1.12
	Control	3.1 (0.8)	—	—	—
Lexical Sophistication	AI-Offload	55.2 (5.4)	7.5	<0.001	0.97
	Control	48.6 (6.1)	—	—	—
Syntactic Complexity	AI-Offload	40.1 (4.8)	6.8	<0.001	0.88
	Control	34.7 (5.2)	—	—	—
Text Cohesion	AI-Offload	4.1 (0.5)	9.8	<0.001	1.27
	Control	3.4 (0.7)	—	—	—

Beyond these rubric dimensions, lexical sophistication (type–token ratio and advanced vocabulary use) increased from M = 48.6 (SD = 6.1) in the control group to M = 55.2 (SD = 5.4) in the AI-offload group, t(238) = 7.5, $p < 0.001$, indicating richer word choice. Syntactic complexity—measured by mean clause length and subordinate clause ratio—rose from M = 34.7 (SD = 5.2) to M = 40.1 (SD = 4.8), t(238) = 6.8, $p < 0.001$, reflecting more complex sentence structures. Finally, text cohesion scores improved from M = 3.4 (SD = 0.7) to M = 4.1 (SD = 0.5), t(238) = 9.8, $p < 0.001$, indicating tighter discourse organization and clearer referential ties.

4.3. Mediation by Cognitive Offloading

Students' self-reported cognitive offloading behavior significantly mediated the relationship between AI-enabled instruction and critical thinking gains. In other words, the structured delegation of routine tasks to AI increased offloading behaviors, which in turn led to higher critical thinking outcomes.

A bias-corrected bootstrap mediation analysis (5,000 samples) was conducted with instructional condition (AI-offload vs. control) as the independent variable, post-test critical thinking score as the dependent variable, and mean cognitive-offloading score (averaged across tasks) as the mediator. As shown in **Table 3**, AI instruction had a significant positive effect on offloading (a = 1.45, SE = 0.12, p < 0.001), and offloading positively predicted critical thinking gains when controlling for condition (b = 4.12, SE = 0.58, $p < 0.001$). The direct effect of instruction on critical thinking remained significant (c′ = 8.67, SE = 1.32, $p < 0.001$), indicating partial mediation. The indirect effect (ab = 5.98) had a 95% bootstrap confidence interval of [3.91, 8.17], not crossing zero, confirming that cognitive offloading significantly mediated the instructional effect.

Table 3. Mediation Analysis Results.

Path	Effect (β or ab)	SE	*p*-Value	95% CI	Interpretation
a: Condition → Offload	1.45	0.12	<0.001	[1.21, 1.69]	AI increased offloading
b: Offload → CT gains	4.12	0.58	<0.001	[2.98, 5.26]	Offloading boosted CT
c: Total effect	14.65	1.28	<0.001	[12.14, 17.16]	Overall effect
c′: Direct effect	8.67	1.32	<0.001	[6.07, 11.27]	Beyond mediation
ab: Indirect effect	5.98	—	<0.001	[3.91, 8.17]	Significant mediation

These results suggest that cognitive offloading serves as a mechanism through which AI-enabled pedagogies enhance higher-order thinking. By offloading lower-order tasks, such as idea generation and surface editing, to AI learners freed up mental resources that they then invested in analysis, evaluation, and reflection. The partial nature of mediation indicates that other factors (e.g., peer discussion, metacognitive prompts) also contributed to critical thinking gains.

5. Discussion

5.1. Balancing Offload and Engagement

This study demonstrates that embedding generative AI within a structured cognitive-offloading framework leads to significant improvements in students' critical thinking skills, surpassing traditional instructional methods. Notably, these enhancements occurred without the "cognitive laziness" often associated with unstructured AI use.

Gerlich identified a negative correlation between frequent, unstructured AI tool usage and critical thinking performance, attributing this decline to excessive delegation of cognitive tasks [21]. Conversely, our scaffolded approach, which incorporates layered prompts, peer-AI co-revision, and reflective journaling, aligns with findings by Yang et al. (2025) [22]. Their study reported that active modification of AI-generated text enhances essay quality and metacognitive engagement. While Yang et al. focused on single-session lab settings, our semester-long intervention in authentic L2 classrooms extends these findings, demonstrating sustained improvements in critical thinking ($\Delta M = +12.3$ points, $\eta^2 = 0.15$).

Furthermore, this study supports the extended mind thesis, suggesting that AI can function as an integrated "cognitive prosthetic" without inducing cognitive atrophy, provided learners engage in intentional reflection on AI outputs [23]. Our mediation analysis reveals that while cognitive offloading significantly contributes to critical thinking gains (indirect effect = 5.98, 95% CI [3.91, 8.17]), direct effects remain significant, indicating that social and metacognitive scaffolds uniquely enhance learning.

By integrating insights from cognitive psychology, L2 writing pedagogy, and AI-education research, this study fills a critical gap by providing controlled, classroom-based evidence that generative AI–enabled cognitive offload instruction can enhance, rather than undermine, critical thinking skills. These findings have significant implications for educators and policymakers aiming to harness AI's efficiencies while preserving intellectual rigor. Specifically, they suggest that balanced offload designs, anchored by metacognitive prompts and collaborative critique, can transform AI from a potential threat into a catalyst for higher-order learning [24,25].

5.2. Pedagogical Implications

This study provides actionable insights for educators aiming to leverage generative AI to enhance, rather than hinder, student learning. By comparing scaffolded AI-offload instruction with traditional pedagogy, the research not only supports previous recommendations but also demonstrates clear improvements in student outcomes, offering a roadmap for effective AI integration in writing classrooms. Unlike unstructured AI workshops that yield modest or even negative effects on critical thinking, embedding layered metacognitive prompts (e.g., "Explain why you accepted or rejected each AI suggestion") produced significant gains in critical thinking ($\eta^2 = .15$). This confirms qualitative claims by Lu et al. that teacher-designed scaffolds are essential for meaningful AI use in EFL contexts and extends them by quantifying the impact in a controlled, semester-long study [26].

Previous case studies [27-30] advocated collaborative critique of AI drafts but lacked statistical validation. Our peer-AI co-revision model not only aligned with those recommendations but yielded significantly higher essay quality metrics—logical coherence ($d = 1.31$) and evidence use ($d = 1.20$)—than control conditions. This demonstrates that social constructivist activities, when combined with AI, amplify both engagement and learning outcomes, improving upon single-session lab findings through evidence of sustained, real-classroom benefits [31-33].

Echoing commentary that AI should augment rather than replace educators' roles, this study's technology fluency training—teaching prompt engineering and cross-platform comparisons—enhanced students' evaluative judgment. While prior research showed that prompt instruction improved surface-level outputs, our data reveal that it also deepened higher-order thinking: students who learned to craft precise prompts demonstrated greater originality (M = 3.9 vs. 3.1) and syntactic complexity (M = 40.1 vs. 34.7). This advances Activity Theory findings by Woo et al., showing that tailored prompt education can be scaled across proficiency levels [34].

Time-saving benefits of AI (30-50% faster drafting) have been documented in cognitive-offloading literature, but critics warn of "cognitive atrophy" without oversight. By demonstrating partial mediation, offloading accounted for an indirect effect (ab) = 5.98 on critical thinking gains. This study shows how efficiency gains can be intentionally redirected into analytic work. This reconciles extended-mind theory with cautionary perspectives like Carr by proving that AI offloading need not diminish deep thinking when accompanied by metacognitive and collaborative scaffolds [35].

6. Conclusions

This study provides empirical evidence that the structured integration of generative AI into L2 writing instruction can significantly enhance learners' critical thinking and essay quality. By employing scaffolded strategies—such as metacognitive prompts, peer-AI co-revision, and reflective journaling—students demonstrated notable improvements in critical thinking and essay quality. These findings align with previous research emphasizing the importance of guided AI use in educational settings. Moreover, the integration of AI tools like ChatGPT with specific guidelines has been found to improve clarity, logical coherence, and evidence use in argumentative writing.

Importantly, this study addresses concerns about AI-iduced cognitive atrophy by demonstrating that, when combined with pedagogical scaffolding, AI can serve as a cognitive extension that enhances rather than diminishes critical engagement. This nuanced understanding contributes to the ongoing discourse on the role of AI in education, suggesting that the key lies not in the technology itself but in how it is integrated into learning environments. Future research should replicate this scaffolded cognitive-offloading model across diverse contexts, examine long-term retention of critical-thinking gains, and explore additional mediators such as motivation and epistemic beliefs. By doing so, the field can build a robust evidence base to guide policy and practice in the AI-augmented classroom.

Author Contributions

Conceptualization, H.H. and P.V.U.L.; methodology, H.H.; software, H.H.; validation, H.H., PP.V.U.L., and C.V.; formal analysis, H.H.; investigation, H.H.; resources, H.H.; data curation, H.H.; writing—original draft preparation, H.H.; writing—review and editing, P.V.U.L. and C.V.; supervision, P.V.U.L. and C.V.; project administration, H.H.. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Ethical review and approval were waived for this study because it involved routine educational practices with minimal risk, and no sensitive personal data were collected.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

The data are not publicly available due to privacy and ethical restrictions but are available from the corresponding author upon reasonable request.

Acknowledgments

The authors would like to thank the participants for their involvement in this study.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Risko, E.F., Gilbert, S.J., 2016. Cognitive Offloading. Trends in Cognitive Sciences. 20(9), 676–688.

[2] Kim, M.K., Kim, N.J., Heidari, A., 2022. Learner

experience in artificial intelligence-scaffolded argumentation. Assessment & Evaluation in Higher Education. 47(8), 1301–1316.

[3] Gerlich, M., 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies. 15(1), 1–28.

[4] Huang, Y., Zhang, L.J., 2022. Facilitating L2 writers' metacognitive strategy use in argumentative writ-ing using a process-genre approach. Frontiers in Psychology. 13, 1–17.

[5] Yang, K., Raković, M., Liang, Z., et al., 2025. Modifying AI, Enhancing Essays: How Active Engagement with Generative AI Boosts Writing Quality. In Proceedings of the 15th International Learning Analytics and Knowledge Conference, Dublin, Ireland, (3–7 March 2025); pp. 568–578.

[6] Yin, Y., Jiahao, Y., 2024. A Proposed Six-step Human–AI Hybrid Argumentative Writing Framework in Higher Education: Enhancing Students' Prompt Engineering Techniques and Critical Thinking Skills. In: O'Dea, X., Ng, D.T.K. (eds.). Effective Practices in AI Literacy Education: Case Studies and Reflections. Emerald Publishing Limited: Leeds, UK. pp. 169–177.

[7] Fan, Y., Tang, L., Le, H., et al., 2025. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology. 56(2), 489–530.

[8] Facione, P.A., 1990. Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). The California Academic Press: Millbrae, CA, USA.

[9] Chu, T., 2020. Cultivating and Assessing Critical Thinkers through Second Language Writing. Frontiers in Educational Research. 3(9), 39–43.

[10] Sato, T., 2022. Assessing critical thinking through L2 argumentative essays: An investigation of relevant and salient criteria from raters' perspectives. Language Testing in Asia. 12(9), 1–19.

[11] Nejad, M.E., Izadpanah, S., Namaziandost, E., et al., 2022. The Mediating Role of Critical Thinking Abilities in the Relationship Between English as a Foreign Language Learners' Writing Performance and Their Language Learning Strategies. Frontiers in Psychology. 13, 1–11.

[12] Shum, S.B., 2024. Generative AI for Critical Analysis: Practical Tools, Cognitive Offloading and Human Agency. In Joint Proceedings of LAK 2024 Workshops, Tokyo, Japan, (18–22 March 2024); CEUR Workshop Proceedings: Aachen, Germany. pp. 205–213.

[13] Fan, Y., Tang, L., Le, H., et al., 2025. Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology. 56(2), 489–530.

[14] Kim, T.W., Tan, Q., 2023. Repurposing Text-Generating AI into a Thought-Provoking Writing Tutor. (arXiv:2304.10543). arXiv. DOI: https://doi.org/10.48550/arXiv.2304.10543

[15] Zou, D., Xie, H., Wang, F.L., 2023. Effects of technology enhanced peer, teacher and self-feedback on students' collaborative writing, critical thinking tendency and engagement in learning. Journal of Computing in Higher Education. 35(1), 166–185.

[16] Woo, D.J., Wang, D., Yung, T., Guo, K., 2024. Effects of a Prompt Engineering Intervention on Undergraduate Students' AI Self-Efficacy, AI Knowledge and Prompt Engineering Ability: A Mixed Methods Study. (arXiv:2408.07302). arXiv. DOI: https://doi.org/10.48550/arXiv.2408.07302

[17] Gupta, A., Shivers-McNair, A. , 2024. "Wayfinding" through the AI wilderness: Mapping rhetorics of ChatGPT prompt writing on X (formerly Twitter) to promote critical AI literacies. Computers and Composition. 74, 102882. DOI: https://doi.org/10.1016/j.compcom.2024.102882

[18] Bouschery, S.G., Blazevic, V., Piller, F.T., 2024. Artificial Intelligence-Augmented Brainstorming: How Humans and Ai Beat Humans Alone. SSRN Electronic Journal. 3(1), 1–47.

[19] Wang, X., Zhang, W., 2024. Generative AI as a Collaborative Companion: Enhancing Peer Feedback in EFL Writing Classes. Education Research and Perspectives (Online). 51, 102–123.

[20] Sachdeva, C., Gilbert, S.J., 2020. Excessive use of reminders: Metacognition and effort-minimisation in cognitive offloading. Consciousness and Cognition. 85, 103024.

[21] Gerlich, M., 2025. AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies. 15(1), 1–28.

[22] Yang, K., Raković, M., Liang, Z., et al., 2025. Modifying AI, Enhancing Essays: How Active Engagement with Generative AI Boosts Writing Quality. In Proceedings of the 15th International Learning Analytics and Knowledge Conference,Dublin, Ireland, (3–7 March 2025). pp. 568–578.

[23] Bozkurt, A., Junhong, X., Lambert, S., et al., 2023. Speculative futures on ChatGPT and generative artificial intelligence (AI): A collective reflection from

the educational landscape. Asian Journal of Distance Education. 18(1), 53–130.

[24] Wang, D., Liu, Y., Jing, X., et al,, 2024. Catalyst for future education: An empirical study on the Impact of artificial intelligence generated content on college students' innovation ability and autonomous learning. Education and Information Technologies. 30(8), 9949–9968. DOI: https://doi.org/10.1007/s10639-024-13209-6

[25] Lu, K., Zhu, J., Pang, F., et al., 2025. Understanding the relationship between colleges students' artificial intelligence literacy and higher order thinking skills using the 3P model: The mediating roles of behavioral engagement and peer interaction. Educational technology research and development. 73(2), 693–716.

[26] Lu R.S., Lin H.C.K., Yang Y.C., et al., 2024. Integrating Urban Mining Concepts Through AI-Generated Storytelling and Visuals: Advancing Sustainability Education in Early Childhood. 16(24), 11304. DOI: https://doi.org/10.3390/su162411304

[27] Emerson, N., 2024. AI-enhanced collaborative story writing in the EFL classroom. Technology in Language Teaching & Learning. 6(3), 1764. DOI: https://doi.org/10.29140/tltl.v6n3.1764

[28] Li, R., 2025. Critiquing ChatGPT Compositions: Collaborative Annotation as an Approach to Enhancing Students' Metalinguistic Awareness of AI-Generated Writing. Thresholds in Education. 48(1), 7–24.

[29] Su, Y., Lin, Y., Lai, C., 2023. Collaborating with ChatGPT in argumentative writing classrooms. Assessing Writing. 57, 100752. DOI: https://doi.org/10.1016/j.asw.2023.100752

[30] Nguyen, A., Hong, Y., Dang, B., et al., 2024. Human-AI collaboration patterns in AI-assisted academic writing. Studies in Higher Education. 49(5), 847–864.

[31] Kartika, S., 2024. Enhancing Writing Proficiency through AI-Powered Feedback: A Quasi-Experimental Study Using Google Gemini. LinguaEducare: Journal of English and Linguistic Studies. 1(2), 83–96.

[32] Li, J., Huang, J., Wu, W., et al., 2024. Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. Humanities and Social Sciences Communications. 11(1), 1–9.

[33] Liu, Z.M., Hwang, G.J., Chen, C.Q., et al., 2024. Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation. Computer Assisted Language Learning. 1–25. DOI: https://doi.org/10.1080/09588221.2024.2389923

[34] Woo, D.J., Guo ,K., and Susanto, H., 2025. Exploring EFL students' prompt engineering in human–AI story writing: An activity theory perspective. Interactive Learning Environments. 33(1), 863–882.

[35] Carr, N., 2020. The Shallows: What the Internet Is Doing to Our Brains (10th anniversary ed.). W. W. Norton & Company: New York, NY, USA.