



Mediating Effects of Essay Writing on the Relationship Between ChatGPT Usage and Critical Thinking in Undergraduate EFL Students: A SmartPLS Approach

Yenni Shintia, Nunung Nurjati*, Endah Yulia Rahayu

Master's Program of English Language Education, Faculty of Teacher Training and Education, Universitas PGRI Adi Buana Surabaya, Surabaya, Indonesia

ARTICLE INFO

Article History

Received : September 16, 2024

1st Revision : October 06, 2024

Accepted : September 29, 2025

Available Online: October 30, 2025

Keywords:

STEM;

Artificial Intelligence; AI ChatGPT;

Critical Thinking; Essay Writing

*Corresponding Author

Email address:

nunung.nurjati@unipasby.ac.id

ABSTRACT

This study investigates whether essay-writing skills mediate the relationship between ChatGPT usage and critical thinking among undergraduate English as a foreign language (EFL) students. Using a quantitative design with partial least squares structural equation modeling (PLS-SEM) in SmartPLS, we analyzed data from 201 students at UINSA Surabaya (167 women; 34 men). PLS-SEM was selected for its suitability in modeling complex mediation with small to medium samples. The measurement model captured essay-writing elements (e.g., clarity, argumentation, coherence) and critical-thinking facets (e.g., inference, reasoning, evaluation). Structural results indicate that ChatGPT usage has a significant, positive association with critical thinking (path coefficient $\beta = 0.401, p < .001$). Essay-writing skills also exhibit a positive effect on critical thinking and function as a meaningful mediator, suggesting that gains in writing quality help translate the use of ChatGPT into higher-order thinking. Diagnostic checks flagged several cross-loadings, indicating potential threats to discriminant validity and construct clarity. Because unaddressed cross-loadings can bias parameter estimates and reduce measurement equivalence, educators and researchers should interpret structural paths with caution and align decisions with theory. Pedagogically, the findings support the guided integration of ChatGPT, emphasizing the critical appraisal of AI output, source evaluation, and iterative drafting, to strengthen students' argumentation and reasoning. We recommend that instructors scaffold reflective prompts and verification routines so that perceived accuracy aligns with actual accuracy, thereby leveraging AI while safeguarding academic rigor.

How to cite: Shintia, Y., Nurjati, N., & Rahayu, E. Y. (2025). Mediating effects of essay writing on the relationship between ChatGPT usage and critical thinking in undergraduate EFL students: A SmartPLS approach. *International Journal of Pedagogy and Teacher Education*, 9(2), 259-269. <https://doi.org/10.20961/ijpte.v9i2.93423>

1. INTRODUCTION

Artificial intelligence (AI) is reshaping contemporary education, and conversational agents, such as ChatGPT, are increasingly integrated into English as a Foreign Language (EFL) instruction. As these tools become integral to coursework, their influence on higher-order outcomes—particularly critical thinking and essay writing—requires rigorous scrutiny. Grounded in constructivist theory, AI-supported interactions can scaffold engagement, prompt reflection, and facilitate iterative meaning-making in writing and thinking (Darwin et al., 2024; Floris et al., 2024). In this view, chatbot prompts, exemplars, and dialogic feedback function as cognitive supports that help learners articulate claims, evaluate evidence, and refine arguments. At the same time, the adoption of AI raises questions about authorship, independence, and the development of durable cognitive strategies in a second-language setting. These competing possibilities motivate a systematic investigation of the conditions under which AI can augment, rather than displace, learners' reasoning and composition processes.

Existing research on AI in education has largely emphasized general applications, leaving the specific pathways through which chatbots shape essay writing and reasoning in EFL contexts underexplored. Early findings suggest that ChatGPT can enhance engagement and personalize practice opportunities; however, concerns persist about authenticity, overreliance, and the potential erosion of learners' autonomous language production (Mhlanga, 2023; Farrokhnia et al., 2024). Critical thinking in EFL is inherently demanding because learners must manage linguistic load while performing higher-order analysis, inference, and evaluation (Paul et al., 2018; Wale & Bishaw, 2020). If appropriately designed, AI-supported activities may lower linguistic barriers

enough to free cognitive resources for reasoning, while still requiring students to justify their claims and critically evaluate sources. Conversely, poorly structured use may encourage superficial acceptance of outputs without interrogation, weakening both language development and critical judgment. A nuanced account is therefore necessary to distinguish between productive and counterproductive uses of AI in academic writing.

Essay writing is a core pedagogical vehicle for cultivating critical thinking because it requires planning, argumentation, evidence use, organization, and revision. In EFL classrooms, structured drafting cycles, analytic rubrics, and targeted feedback align writing subskills with cognitive processes central to reasoning, including clarity of purpose, logical structure, and originality (Hughes, 2014; Fithriani, 2021). ChatGPT can contribute to these processes by generating alternatives, modeling language patterns, and prompting elaboration, thereby stimulating metacognitive monitoring and self-explanation. The novelty of the present study lies in integrating essay writing, critical thinking, and chatbot use within a single model to test how these elements interact in a real instructional context. Using structural equation modeling (SEM), we estimate direct and indirect relations among ChatGPT use, essay writing quality, and critical thinking performance. Preliminary estimates indicate a substantial positive effect of ChatGPT on essay writing ($\beta = 0.601$), a moderate direct effect on critical thinking ($\beta = 0.183$), and a moderate effect of essay writing on critical thinking ($\beta = 0.136$), suggesting plausible mediation through writing. These values motivate a confirmatory test on a larger sample of EFL undergraduate students.

Despite mounting interest, a clear research gap remains regarding the mediating role of essay writing in the relationship between ChatGPT use and critical thinking among EFL undergraduates. The present study addresses this gap by modeling ChatGPT use as an exogenous predictor, essay writing as a mediator, and critical thinking as the outcome, with measurement attentive to both linguistic and cognitive dimensions. Specifically, we examine whether the use of ChatGPT exerts a positive direct effect on critical thinking and essay writing, whether essay writing has a positive direct effect on critical thinking, and whether essay writing mediates the effect of ChatGPT use on critical thinking. The design acknowledges both benefits and risks by embedding tasks that require justification, source evaluation, and iterative revision, ensuring that AI assistance complements rather than supplants student cognition (Floris et al., 2024). By articulating these pathways, the study contributes theory-driven evidence on the responsible integration of AI to support complex cognitive skills in EFL writing. Ultimately, the findings aim to inform instructional design, assessment practices, and policy guidelines for the effective and ethical use of AI in higher education language learning.

2. MATERIAL AND METHOD

Research Design and Setting

This study adopted a quantitative, cross-sectional survey design to investigate the effects of ChatGPT use (X) on undergraduate EFL students' critical thinking (Y), with essay writing (Z) as a mediator. Structural relations were estimated using Partial Least Squares Structural Equation Modeling (PLS-SEM) in SmartPLS, a software chosen for its suitability in modeling complex relationships with mediation (Goodwin & Liu, 2023). To corroborate construct validity, we additionally inspected measurement properties via confirmatory factor analysis (CFA) in AMOS 26.0. The study was conducted at Universitas Islam Negeri Sunan Ampel (UINSA), Surabaya, Indonesia. Data were collected over approximately two months following the approval of the research proposal, providing a snapshot of current AI integration in EFL coursework.

Participants and Sampling

Participants were undergraduate students enrolled in the English Education program at UINSA. Simple random sampling was applied; the minimum sample size was estimated using Slovin's formula. The final sample comprised $N = 201$ students (167 female, 34 male; 83% female, 17% male). Regarding age, 35 students (17%) were ≤ 19 years (semester 3) and 49 students (24%) were 19–20 years (semester 5); the remainder were above 20. Inclusion criteria were active enrollment and ongoing participation in English academic writing tasks. Participation was voluntary, and informed consent was obtained prior to data collection.

Instruments and Pre-Survey Development

A structured questionnaire of 31 items was administered: items 1–7 captured demographics; items 8–31 operationalized the three latent constructs—ChatGPT use (perceived accuracy, feature availability, efficiency, adaptability, data availability), essay writing (clarity of purpose, logical structure, analytical thinking, evidence

and scholarship, coherence and flow, conciseness and clarity, introduction, grammar/mechanics, originality, counterargument), and critical thinking (questioning, analytical thinking, evidence evaluation, inference, problem solving, creativity, open-mindedness, effective communication). All items used a five-point Likert scale.

Table 1. Likert scale mapping

Response option	Value	Range
Strongly Disagree	1	1.00–1.80
Disagree	2	1.81–2.60
Less Agree	3	2.61–3.40
Agree	4	3.41–4.20
Strongly Agree	5	4.21–5.00

A brief pre-survey informed instrument refinement (e.g., wording and coverage of functions such as idea generation, grammar/style, research assistance, argument structuring, and revision). Descriptive agreement rates appear in Table 2.

Table 2. Blueprint and Indicators from the Pre-Survey (AI–Critical Thinking–Essay Writing)

Construct/Relation	Indicator	What it measures (concise)	% Agree*
AI → Critical Thinking	Idea support	AI helps generate, refine, and organize ideas; improves clarity	63.3
	Grammar & style	AI improves clarity, readability, and consistency	66.6
	Research & information	AI helps find relevant, reliable sources for decision-making	66.6
	Argument structuring	AI supports clear, persuasive argument organization	60.0
	Feedback & revision	AI aids iterative improvement and alignment with goals	50.0
Essay Writing → Critical Thinking	Information analysis	Analyzing sources/patterns to inform decisions and problem-solving	66.6
	Evidence evaluation	Judging the validity/relevance of evidence to support claims	50.0
	Logical reasoning	Logical flow, argumentation, and avoidance of fallacies	50.0
AI → Essay Writing	Outlining & structure	ChatGPT helps create outlines and structure essays	60.0
	Utilization awareness	Awareness/experience using ChatGPT for learning tasks	70.0
	Research support	Summarizing articles and extracting key points	80.0
	Planning & organization	Clarifying main ideas and organizing essay sections	89.0

Procedure and Data Collection

Following institutional approvals, the survey was administered during regular coursework. Students completed the questionnaire independently, with instructions emphasizing honest self-report, anonymity, and confidentiality. The pre-survey phase elicited perceptions of ChatGPT's functions in idea support, grammar/style, information seeking, argument structuring, and revision; the resulting agreement rates guided minor refinements to item wording and coverage. (Ethical compliance. The protocol adhered to UINSA's Institutional Review Board standards. Students received study information and provided informed consent before participation. Anonymity and confidentiality were guaranteed; participation was voluntary with the option to withdraw at any time. Data were stored securely and analyzed in a manner that ensured confidentiality and maintained anonymity.

Table 3. Survey Instrument for AI Use, Critical Thinking, and Essay Writing: Constructs, Indicators, and % Agreement (N = 201)

XP Code	AI Use Indicator	% Agree	YP Code	Critical Thinking Indicator	% Agree	ZP Code	Essay Writing Indicator	% Agree
XP1	Perceived accuracy	77.6	YP1	Questioning	80.1	ZP1	Thesis clarity	79.6
XP2	Answer satisfaction	75.1	YP2	Analytical thinking	77.1	ZP2	Topic sentences/structure	75.1
XP3	Usefulness/detail	77.1	YP3	Evidence evaluation	77.6	ZP3	Own reasoning	80.1
XP4	Adapts to user style	79.6	YP4	Inference	79.1	ZP4	Citations & scholarship	81.6
XP5	Adaptability to needs	74.1	YP5	Problem solving	77.6	ZP5	Coherence & flow	84.6
XP6	Current info accuracy	76.1	YP6	Creativity	87.6	ZP6	Conciseness & clarity	79.6
			YP7	Open-mindedness	81.1	ZP7	Brainstorming (intro)	80.1
			YP8	Communication	81.6	ZP8	Mechanics & citation	81.1
						ZP9	Originality (no plagiarism)	83.6
						ZP10	Counterargument use	82.6

Data Analysis and Model Evaluation

Analyses were conducted in SmartPLS (PLS-SEM) with 5,000-sample bootstrapping for inference. The measurement model was evaluated using indicator loadings, Cronbach's alpha, composite reliability (targets ≥ 0.70), Average Variance Extracted ($AVE \geq 0.50$), and discriminant validity assessed via Fornell–Larcker and HTMT. A supplemental CFA in AMOS 26.0 corroborated convergent and discriminant validity. The structural model was assessed through path coefficients, R^2 , f^2 , and Q^2 . PLSpredict indicated out-of-sample performance with RMSE = 0.812 for both critical thinking and essay writing, MAE = 0.584 (critical thinking) and 0.618 (essay writing). Reliability was satisfactory, with all alpha and composite reliability values exceeding 0.70, and the overall fit was high (GoF = 0.5002). Hypothesis tests examined (i) direct effects of ChatGPT use on essay writing and critical thinking, (ii) the direct effect of essay writing on critical thinking, and (iii) the mediating role of essay writing in the ChatGPT $\rightarrow$ critical thinking pathway.

3. RESULTS

Measurement Model

The measurement model was assessed using established reliability and validity criteria. Internal consistency was examined through composite reliability (CR) and Cronbach's alpha (α) to determine whether items within each construct coherently measure the same latent concept. CR provides a loading-weighted estimate of reliability, whereas α assumes equal loadings across items; considering both indices offers a more robust appraisal of consistency. In line with common practice, values of 0.70 or higher on these indices indicate acceptable reliability and suggest dependable measurement. Convergent validity was appraised by inspecting outer loadings, which reflect the strength of the relationship between each indicator and its construct. Loadings ≥ 0.70 are desirable because they imply that at least 49% of the indicator's variance is captured by the latent construct, thereby supporting the adequacy of the measurement model.

Discriminant validity was evaluated using complementary criteria to ensure that constructs are empirically distinct. First, the Fornell–Larcker criterion was applied by comparing the square root of each construct's average variance extracted (AVE) with its correlations with other constructs; superiority of the square

root of AVE indicates that a construct shares more variance with its own indicators than with other constructs. Second, the Heterotrait–Monotrait ratio (HTMT) was inspected to contrast between-construct correlations with within-construct correlations; HTMT values below 0.90 denote acceptable discriminant validity. Employing both procedures reduces the risk of construct overlap and strengthens the interpretability of structural relations. Together with the outer loading evidence, these tests provide a coherent validation narrative. Overall, the combined results indicate that the indicators are reliable, the constructs converge appropriately, and the latent variables are sufficiently distinct for subsequent structural analysis.

Structural Model

The structural (inner) model specifies the hypothesized relations among ChatGPT use, essay writing, and critical thinking, and it is diagrammed in Figure 1 (Structural Model) to make the directional links and the mediating pathway explicit. Estimates were obtained with PLS-SEM using 5,000 bootstrap resamples and two-tailed tests at $\alpha = 0.05$, which supports stable inference for both direct and indirect effects. We report standardized coefficients (β), standard deviations, t -statistics, and p -values to ensure transparent evaluation of effect magnitudes and significance. This specification enables simultaneous estimation of multiple paths while accounting for measurement error inherent in latent constructs. The diagram in Figure 1 also clarifies that essay writing is modeled as a proximal academic process through which ChatGPT may influence higher-order cognition. Interpreting the model, therefore, requires attending to both the direct route from ChatGPT to critical thinking and the indirect route via writing quality. Overall, this configuration aligns with the theoretical expectation that AI-supported composing can facilitate reasoning in EFL settings.

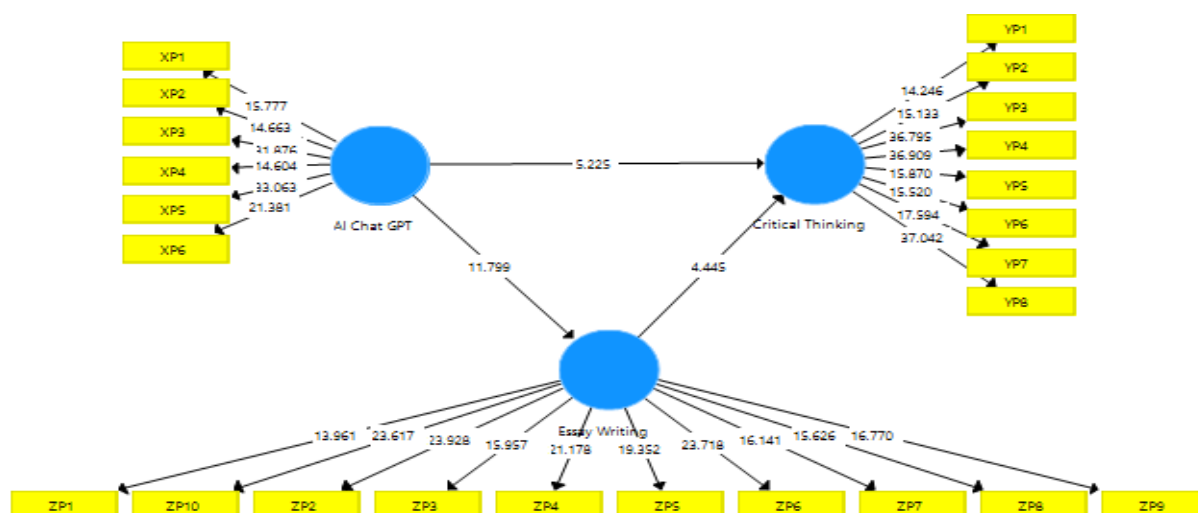


Figure 1. Structural Model (inner model)

Direct path estimates are uniformly positive and statistically significant, as summarized in Table 4 for ease of reference. The effect from ChatGPT $\rightarrow$ Essay Writing is strong ($\beta = 0.613$, $t = 11.799$, $p < 0.001$), indicating substantial improvements in organization and argumentative structure associated with greater ChatGPT use. The path ChatGPT $\rightarrow$ Critical Thinking is moderate ($\beta = 0.401$, $t = 5.225$, $p < 0.001$), suggesting a meaningful direct contribution of AI to students' reasoning beyond its influence on writing. The path Essay Writing $\rightarrow$ Critical Thinking is also moderate ($\beta = 0.346$, $t = 4.445$, $p < 0.001$), consistent with the role of structured composition in analysis, evaluation, and argument construction. Taken together, these coefficients support H1–H3 and align with the directional assumptions depicted in Figure 1. The convergence of statistical significance across all three links strengthens confidence in the theorized mechanism. The coherence between the numeric evidence in Table 4 and the pathways in Figure 1 enhances the internal consistency of the results.

Mediation analysis further clarifies the mechanism by which ChatGPT relates to critical thinking, and the results are presented in Table 5. The indirect path ChatGPT $\rightarrow$ Essay Writing $\rightarrow$ Critical Thinking is significant and positive ($\beta_{\text{indirect}} = 0.212$, $t = 4.112$, $p < 0.001$), indicating partial mediation rather than a purely direct influence. In practical terms, part of AI's benefit for critical thinking is realized through improvements in essay

organization, evidence use, and argumentative coherence. This pattern complements the strong direct effect on writing and the moderate direct effect on reasoning, reinforcing a writing-centered route to cognitive development. Pedagogically, the findings recommend integrating ChatGPT within guided drafting, source evaluation, counterargument construction, and iterative revision cycles. The mediated pathway visualized in Figure 1 is thus empirically supported by the statistics reported in Table 5. Finally, the consistency between direct and indirect effects suggests that responsible, task-embedded AI use can be leveraged to strengthen higher-order thinking in EFL contexts.

Table 4. The Hypothesis of Path Coefficients in Specific Direct Effects:

	Path Coefficient/ Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (O/STD EV)	P Values < 0,05
AI Chat GPT -> Critical Thinking	0,401	0,404	0,077	5,225	0,000
AI Chat GPT -> Essay Writing	0,613	0,621	0,052	11,799	0,000
Essay Writing > Critical Thinking	0,346	0,351	0,078	4,445	0,000

Table 5. Path Coefficients in Specific Indirect Effects

	Path Coefficients/ Original Sample (O)	Sample Mean (M)	Standard Deviation (STDEV)	T Statistics (O/ST DEV)	P Values
AI Chat GPT → Essay Writing → Critical Thinking	0,212	0,218	0,052	4,112	0,000

Model Fit (Goodness of Fit)

The overall Goodness of Fit (GoF) for the model linking ChatGPT use, essay writing, and critical thinking is 0.5002, indicating a moderate to high level of fit for an exploratory PLS-SEM study. Component values show communalities/AVE of 0.636 for Critical Thinking and 0.576 for Essay Writing, alongside R^2 values of 0.450 and 0.375, respectively, reflecting meaningful explained variance in both endogenous constructs. The aggregated figures (AVE $\approx$ 0.606, mean $R^2 \approx$ 0.413) yield the reported GoF of approximately 0.50, which supports the adequacy of the measurement–structural specification. Values closer to 1 indicate a better fit, and a GoF around 0.50 is generally interpreted as acceptable in early-phase, model-building research. In this context, the fit statistics complement the significant paths observed in the structural model, suggesting that the theorized relations are empirically defensible. Although GoF does not substitute for detailed reliability and validity checks, it provides a compact, global snapshot that the model captures a substantial share of the observed covariance structure. Taken together, these indicators justify proceeding to predictive evaluation and substantive interpretation.

Table 6. Result of the Goodness of Fit (GoF) index

	Communality score	R Square	GoF Index
Critical Thinking	0,636	0,450	
Essay Writing	0,576	0,375	
Total AVE communality	0,606	0,413	0,5002187

While the GoF is encouraging, it should be read in conjunction with construct-level evidence to avoid overgeneralization. The $R^2 = 0.450$ for critical thinking implies that nearly half of its variance is accounted for by

the use of ChatGPT and essay writing, which is sizable for complex cognitive outcomes. The $R^2 = 0.375$ for essay writing likewise indicates moderate explanatory power, consistent with a multifactorial skill that integrates language, organization, and argumentation. The relatively high communalities/AVE suggest that indicators load strongly on their intended constructs, reinforcing convergent validity established earlier. These results align with the strong direct effect of ChatGPT on essay writing and the moderate effects on and from critical thinking reported in the structural model. From a practical standpoint, the combination of an acceptable fit and meaningful R^2 values indicates that the model is useful for informed instructional decision-making and future refinement. Nonetheless, incremental improvements—such as expanding indicators for originality or source evaluation—could raise common variance and sharpen discriminant boundaries in subsequent replications. Overall, the GoF profile supports the credibility of the model while leaving room for theoretically motivated enhancements.

Predictive Performance and Synthesis of Findings

Predictive relevance was assessed with PLSpredict, and results are summarized in Table 7 (LV Prediction Summary) to provide an out-of-sample perspective. The Q^2_{predict} values are 0.361 for Critical Thinking and 0.354 for Essay Writing, both of which are greater than 0, indicating that the model has meaningful predictive capability for each latent variable. Error metrics were comparable across outcomes: RMSE $\approx$ 0.812 for both constructs, MAE = 0.584 for critical thinking, and MAE = 0.618 for essay writing, suggesting stable prediction error magnitudes. These patterns imply slightly stronger predictive performance for critical thinking relative to essay writing, even though writing serves as the mediator in the structural chain. In predictive terms, the model generalizes beyond the estimation sample with acceptable accuracy for an educational setting characterized by heterogeneous learning profiles. The alignment between positive Q^2 values and moderate R^2 estimates strengthens confidence that the model is not merely descriptive but also forward-looking. As such, PLSpredict complements the global fit indices by addressing the practical question of how well the model forecasts unseen responses.

Table 7. Prediction Summary

	Q^2 Predict	RMSE	MAE
Critical Thinking	0,360559108	0,811932316	0,584310367
Essay Writing	0,354306148	0,812127976	0,618453841

Synthesizing the structural and predictive evidence yields clear substantive conclusions. First, ChatGPT $\rightarrow$ Essay Writing is strong ($\beta = 0.613$, $t = 11.799$, $p < 0.001$), indicating that AI-supported composing substantially improves organization, coherence, and argumentative development. Second, ChatGPT $\rightarrow$ Critical Thinking is moderate ($\beta = 0.401$, $t = 5.225$, $p < 0.001$), demonstrating a meaningful direct contribution of AI use to reasoning quality beyond its effect on writing. Third, Essay Writing $\rightarrow$ Critical Thinking is moderate ($\beta = 0.346$, $t = 4.445$, $p < 0.001$), underscoring the role of structured composition in analysis, evaluation, and inference. Fourth, the indirect effect ChatGPT $\rightarrow$ Essay Writing $\rightarrow$ Critical Thinking is positive and significant ($\beta_{\text{indirect}} = 0.212$, $t = 4.112$, $p < 0.001$), confirming partial mediation and identifying writing as a proximal mechanism for cognitive gains. Fifth, the GoF = 0.5002 and $Q^2_{\text{predict}} > 0$ for both outcomes jointly indicate a model that fits reasonably well and predicts usefully in an applied EFL context. Sixth, these findings recommend writing-centered AI integration—including guided outlining, source appraisal, counterargument, and iterative revision—to maximize gains in both writing skill and higher-order thinking. Finally, because the model demonstrates both explanatory and predictive adequacy, it provides a credible foundation for instructional design, program evaluation, and future confirmatory studies that extend the indicator set and test robustness across cohorts.

4. DISCUSSION

Impact of ChatGPT on Critical Thinking

The present study demonstrates a significant and positive direct association between ChatGPT use and students' critical thinking in an EFL context. This finding supports the view that AI-enabled dialogue can cultivate engagement, self-monitoring, and iterative reflection, which are foundational to higher-order reasoning. By

enabling rapid question generation, immediate response appraisal, and cycles of revision, ChatGPT creates conditions that approximate reflective practice within a manageable time frame. Such conditions allow learners to test conjectures, examine counterpositions, and refine claims with greater frequency than typical classroom schedules permit. The result is not merely an acceleration of language production, but an expansion of opportunities for analytical processing. Prior literature has converged on similar patterns, noting that conversational agents can act as flexible cognitive tools that adapt to task demands and stimulate interactional competence (Mahapatra, 2024; Maghamil & Sieras, 2024). At the same time, the present evidence suggests that gains in reasoning quality are not uniform and may depend on how prompts structure analysis, evidence, and justification. These nuances underscore the importance of task design that prioritizes argument quality over surface correctness.

The literature also cautions that over-reliance on AI may hinder the development of independent monitoring and in-depth analysis when feedback is accepted uncritically. Mixed-method studies report that students perceive clear improvements in language clarity and organization but still benefit from targeted human guidance for advanced evaluative judgment and synthesis (Prasetya & Syarif, 2023). Our findings are consistent with that caution, because statistical significance does not preclude variation in individual strategy use or differences in metacognitive uptake. Put differently, the presence of a positive average effect does not guarantee optimal learning for all learners or tasks. Educators, therefore, need to scaffold interpretation of AI outputs, asking students to justify revisions, identify unstated assumptions, and locate gaps in reasoning. These metacognitive moves transform AI support from an answer provider into a catalyst for inquiry. When positioned this way, ChatGPT serves as a rehearsal space for analysis rather than a substitute for it. The overarching implication is that improvements in critical thinking are most plausible when AI use is embedded within purposefully sequenced analytical tasks.

The Mediating Role of Essay Writing

The study finds that essay writing functions as a meaningful conduit through which ChatGPT use relates to critical thinking, indicating partial mediation. This mechanism is pedagogically plausible because writing compels learners to articulate a stance, marshal evidence, and structure claims in a logically coherent sequence. ChatGPT can enhance these processes by providing instantaneous suggestions on thesis clarity, topic-sentence focus, and transitions, which are recurring bottlenecks in EFL composition (Mhlanga, 2023). Such support does not merely correct sentences; it reorganizes conceptual flow so that arguments unfold with clearer warrants and better signposting. As students iterate between drafting and feedback, they externalize reasoning steps and become more attentive to coherence relations across paragraphs. Over time, this cycle can consolidate habits of analysis and explanation that generalize beyond a single assignment. Importantly, mediation suggests that a portion of the cognitive benefit stems from improved writing quality, rather than solely from direct AI-reasoning interaction. This pattern highlights the importance of writing-centered curricula in promoting cognitive development in second-language settings.

Beyond structure, ChatGPT can assist with lexical precision, syntactic variety, and genre-appropriate phrasing, which jointly elevate the readability and persuasiveness of essays. Studies suggest that such tailored feedback promotes self-reflection and sustained motivation by enabling learners to diagnose and repair errors independently over time (Nizzolino, 2024). However, the same affordances can become liabilities if students substitute AI text for their own reasoning or bypass source evaluation. To mitigate this risk, instructors should require them to maintain transparent revision logs that document the rationale for accepting or rejecting AI suggestions. They should also integrate checkpoints for evidence appraisal, including credibility checks and a balance of paraphrasing and quotations. These measures encourage active authorship and discourage passive uptake, thereby preserving the cognitive value of writing tasks. When implemented thoughtfully, AI-supported drafting becomes a scaffold for analysis rather than a shortcut around it. In that configuration, the mediating role of writing is not incidental but central to the cultivation of higher-order thinking.

Practical Implications for EFL Educators

The findings recommend integrating ChatGPT as a form of formative support within writing-intensive instruction, rather than as a stand-alone solution. Teachers can assign staged drafts in which students first generate outlines, then expand to body paragraphs, and finally revise based on AI-informed, rubric-aligned

prompts. Each stage should include analytic checkpoints that ask learners to justify claims, connect evidence to warrants, and craft counterarguments. This sequencing keeps the cognitive load productive by aligning language work with reasoning goals. To maintain authorship, students should annotate AI-assisted revisions, explaining why changes improve clarity, logic, or evidential support. Such annotation practices foster metacognitive awareness and help instructors differentiate between superficial edits and substantive improvements. In parallel, peer review can be organized around criteria that mirror the analytic rubric, thereby triangulating feedback sources. When these elements work together, AI becomes a partner in learning rather than a proxy for it.

ChatGPT can also be used to stimulate critical dialogue, but prompts must be crafted to elicit evaluation rather than agreement. Instructors might require students to critique AI-generated outlines for hidden assumptions, missing counterevidence, or weak causal links. They can require students to propose revisions that strengthen validity or reliability claims, thereby operationalizing standards of argument quality. Ethical and procedural guidelines should accompany these activities, including expectations for citation, disclosure of AI assistance, and protection of personal data (Teng, 2024). To prevent skill atrophy, classes should alternate AI-enhanced tasks with AI-off assessments that capture independent performance. This balance suggests that technology serves as a scaffold, not a replacement, for linguistic and cognitive work. Finally, professional development should equip educators with effective engineering strategies, design skills for rubrics, and literacy in analytics, so that they can interpret model outputs responsibly. In this ecosystem, ChatGPT augments communicative language teaching by widening opportunities for practice while preserving the integrity of intellectual work.

Future Research Directions

Future investigations should prioritize longitudinal and multi-cohort designs to test durability and transfer of AI-assisted gains in reasoning and writing. Long-term exposure may yield different trajectories than short interventions, and only repeated measurements can reveal consolidation or decay of effects (Montgomery, 2023). Researchers should include AI-off baselines and delayed posttests to separate immediate assistance from enduring competence. Mixed-method components—such as think-aloud protocols, revision logs, and classroom observations—can illuminate how students internalize feedback into strategies. Measurement should extend beyond global test scores to include fine-grained indicators of warranting, counterargument, and evidence integration. Where feasible, studies should incorporate learning analytics to model the relationship between prompt types, feedback timing, and revision frequency, as well as their outcomes. Attention to equity is also essential, as access to devices, bandwidth, and prior digital literacy can moderate the effects of treatment. Collectively, these choices will enhance the accuracy of causal inference and improve the interpretability of results across various settings.

Comparative and differential-effects research is likewise needed to position ChatGPT among adjacent tools and populations. Randomized or quasi-experimental studies could compare ChatGPT with grammar-focused assistants, domain-specific tutors, or human feedback under matched time-on-task conditions (Barton et al., 2024). Researchers should also examine heterogeneity by proficiency level, disciplinary writing demands, and motivational profiles, as students in STEM and humanities may leverage AI differently (Lo et al., 2024). Cross-institutional replications can test the robustness of findings across different curricula, instructor expertise, and assessment cultures. Instrument development also deserves attention, including the development of validated rubrics for originality, source quality, and reasoning depth that are sensitive to AI-influenced writing. Ethical dimensions—such as authorship transparency and data privacy—should be embedded as outcomes, not just constraints, so that pedagogy aligns with the responsible use of AI. Ultimately, a coordinated agenda that blends methodological rigor with ecological validity will clarify when, for whom, and under what conditions AI-enabled writing most effectively advances critical thinking in EFL education.

5. CONCLUSION

This study advances understanding of how AI language models—specifically ChatGPT—shape learning in EFL contexts by jointly strengthening essay writing and critical thinking. The results indicate a meaningful direct effect of ChatGPT on students' reasoning, while also revealing a significant indirect pathway through improved writing performance. Together, these effects suggest that AI-supported composing creates productive conditions for analysis, inference, and argumentation rather than merely accelerating text production. Pedagogically, the

evidence supports positioning ChatGPT as a formative scaffold embedded in writing-intensive tasks that demand explanation, evidence appraisal, and revision. Such integration should be guided by explicit prompts, analytic rubrics, and transparency regarding AI assistance to maintain authorship and foster deeper metacognitive engagement. At the same time, safeguards against over-reliance are essential, including AI-off assessments and reflective justification of AI-informed revisions. Programmatically, the findings recommend professional development for instructors in prompt design, feedback orchestration, and the ethical use of technology to align it with learning goals. Methodologically, the conclusions are reinforced by reliable measurement and an adequately fitting structural model, which together provide confidence in the inferences drawn.

The practical implications suggest a balanced approach: AI should augment rather than replace human instruction, thereby amplifying opportunities for iterative drafting, targeted feedback, and authentic argument construction. Curriculum designers can leverage ChatGPT to support outlining, source evaluation, counterargument, and cohesion work, while preserving space for teacher judgment and peer review. Policy guidelines should codify disclosure practices, data privacy protections, and assessment standards to ensure integrity and fairness in AI-mediated coursework. Future research should prioritize longitudinal and mixed-method designs to test durability, transfer, and mechanism, including delayed posttests and analysis of revision trails. Moderation by proficiency, genre, and learner profile warrants careful attention, as benefits may vary across subgroups and discourse demands. Comparative trials that contrast ChatGPT with other support tools can clarify the unique and overlapping effects on writing quality and reasoning depth. Instrument development should continue, with validated indicators for originality, evidential reasoning, and rhetorical control that are sensitive to AI-influenced writing. Taken together, these steps will establish an evidence-based pathway for utilizing AI to achieve the dual aims of language proficiency and higher-order thinking in EFL education.

6. REFERENCES

- Barton, H., Vlachopoulos, D., Benjamin, A., & Opuni, J. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence*, 6, 100198. <https://doi.org/10.1016/j.caeai.2023.100198>
- Darwin, Rusdin, D., Mukminatien, N., Suryati, N., Laksmi, E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, 11(1), 2290342. <https://doi.org/10.1080/2331186X.2023.2290342>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, 61(3), 460–474. <https://doi.org/10.1080/14703297.2023.2195846>
- Fithriani, R. (2021). The utilization of mobile-assisted gamification for vocabulary learning: Its efficacy and perceived benefits. *Computer Assisted Language Learning Electronic Journal*, 22(3), 146–163.
- Floris, F. D., Widiati, U., Renandya, W. A., & Basthomi, Y. (2024). Artificial intelligence in English language teaching: Fostering joint enterprise in online communities. *JEES (Journal of English Educators Society)*, 9(1), 12–21. <https://doi.org/10.21070/jees.v9i1.1825>
- Goodwin, M., & Liu, Y. (2023). Investigating SmartPLS in education research: Applications in language skills. *Journal of Educational Research*, 116(4), 409–419. <https://doi.org/10.1080/00220671.2023.1324434>
- Hughes, J. (2014). *Critical thinking in the language classroom*. ELI Publishing.
- Lo, C. K., Hew, K. F., & Jong, M. S. (2024). The influence of ChatGPT on student engagement: A systematic review and future research agenda. *Computers & Education*, 219, 105100. <https://doi.org/10.1016/j.compedu.2024.105100>
- Maghamil, M., & Sieras, S. (2024). Impact of ChatGPT on the academic writing quality of senior high school students. *Journal of English Language Teaching and Applied Linguistics*, 6(2), 115–128. <https://doi.org/10.32996/jeltal.2024.6.2.14>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed-methods intervention study. *Smart Learning Environments*, 11(9), 1–18. <https://doi.org/10.1186/s40561-024-00295-9>

-
- Mhlanga, D. (2023). OpenAI in education: The responsible and ethical use of ChatGPT towards lifelong learning. In *Responsible AI in higher education* (pp. 387–409). Springer. https://doi.org/10.1007/978-3-031-37776-1_17
- Montgomery, D. P. (2023). “This study is not without its limitations”: Acknowledging limitations and recommending future research in applied linguistics research articles. *Journal of English for Academic Purposes*, 65, 101291. <https://doi.org/10.1016/j.jeap.2023.101291>
- Nizzolino, S. (2024). Artificial Intelligence in Language Teaching: Using ChatGPT to assist teachers of English as a Foreign Language. *Form@re - Open Journal Per La Formazione in Rete*, 24(1), 242–261. <https://doi.org/10.36253/form-15681>
- Paul, R., Elder, L., & Bartell, T. (2018). A brief history of the idea of critical thinking. Kwelanga Training (Pty) Ltd.
- Prasetya, R. E., & Syarif, A. (2023). ChatGPT as a tool for language development: Investigating its impact on proficiency and self-evaluation accuracy in Indonesian higher education. *VELES (Voices of English Language Education Society)*, 7(3), 402–415. <https://doi.org/10.29408/veles.v7i3.19303>
- Teng, M. F. (2024). “ChatGPT is the companion, not enemies”: EFL learners’ perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- Wale, B. D., & Bishaw, K. S. (2020). Effects of using inquiry-based learning on EFL students’ critical thinking skills. *Asian-Pacific Journal of Second and Foreign Language Education*, 5(1), 1–17. <https://doi.org/10.1186/s40862-020-00090-2>