

Integrating ChatGPT within Genre-Based Pedagogy to Strengthen Critical Thinking in ESP Academic Writing

Eliasanti Agustina^{1*}; Ahmad Munir²; Slamet Setiawan³

¹Universitas Negeri Surabaya

²Universitas Negeri Surabaya

³Universitas Negeri Surabaya

¹ORCID: 0009-0009-2164-9574

²ORCID: 0000-0003-0606-3590

³ORCID: 0000-0003-4143-8757

Corresponding Author: Eliasanti Agustina^{1*}

Publication Date: 2026/10/22

Abstract: This study investigates the effectiveness of integrating ChatGPT within genre-based pedagogy to strengthen critical thinking in ESP (English for Specific Purposes) academic writing among students in East Java higher education institutions. Using a mixed-methods sequential explanatory design, the research involved 142 students from six private universities. Quantitative data were collected through pre-test and post-test argumentative essays, analyzed using paired sample t-tests, while qualitative data were obtained through classroom observations, semi-structured interviews, and document analysis. Results demonstrated statistically significant improvements ($p < 0.05$) in three critical thinking dimensions: argument evaluation ($M = 2.71$ to $M = 3.82$), problem analysis ($M = 2.64$ to $M = 3.75$), and decision justification ($M = 2.58$ to $M = 3.69$). Qualitative findings revealed that students perceived ChatGPT as a dialogic cognitive partner that facilitated idea development, argument evaluation, and reflective revision throughout the teaching-learning cycle. However, students also recognized the need for metacognitive regulation to prevent over-reliance on technology. This research contributes an integrated model combining AI-assisted writing tools, structured pedagogical design, and explicit critical thinking instruction, offering practical implications for digital transformation in Indonesian higher education.

Keywords: ChatGPT, Genre Based Pedagogy, Critical Thinking, Academic Writing, English for Specific Purposes, Artificial Intelligence.

How to Cite: Eliasanti Agustina; Ahmad Munir; Slamet Setiawan (2026) Integrating ChatGPT within Genre-Based Pedagogy to Strengthen Critical Thinking in ESP Academic Writing. *International Journal of Innovative Science and Research Technology*, 11(10), 1306-1314. <https://doi.org/10.38124/ijisrt/26oct016>

I. INTRODUCTION

➤ Digital Transformation and AI in Academic Writing Instruction

Higher education in the 21st century experiences systematic digital transformation through artificial intelligence (AI) integration in academic practices. Generative AI models such as ChatGPT have fundamentally altered academic writing from linear, individual-based practice to dialogical, human-machine interactive processes. In writing instruction, AI functions not merely as a grammar correction tool but as a cognitive support system providing argumentative feedback, suggesting idea restructuring, and facilitating metacognitive reflection. Empirical studies

demonstrate that students leverage ChatGPT to enhance essay coherence, clarify thesis statements, and enrich idea elaboration. However, concerns regarding academic integrity and cognitive dependency persist when AI usage lacks systematic pedagogical regulation. Therefore, AI integration must be understood as structured instructional design, not merely technology adoption.

➤ Critical Thinking as Core Competency

Higher education globally emphasizes critical thinking as a core 21st-century graduate competency. In academic writing, critical thinking extends beyond opinion expression; it encompasses argument evaluation, problem analysis, assumption testing, and evidence-based decision

justification. The Watson-Glaser framework identifies five key dimensions: inference, assumption recognition, deduction, interpretation, and argument evaluation. Within argumentative writing, these dimensions manifest through logical thesis construction, coherent reasoning development, credible evidence integration, and counterargument consideration. Research indicates students often construct claims effectively but struggle with evaluating argument strength and critically elaborating evidence, suggesting writing instruction must explicitly integrate critical thinking frameworks.

➤ *Genre-Based Pedagogy and Systemic Functional Linguistics*

Systemic Functional Linguistics (SFL) provides theoretical grounding for understanding how language realizes meaning in argumentative texts. SFL views language as a social semiotic system with three metafunctions: ideational (representing cause-effect and logical structure), interpersonal (reflecting writer stance through modality and evaluation), and textual (organizing information through theme-rheme progression and lexical cohesion). Argumentative quality depends not only on content accuracy but also on how linguistic structures support logical strength. Genre-based pedagogy operationalizes SFL through a Teaching-Learning Cycle (TLC): modeling (analyzing model texts), joint construction (instructor-student collaborative text building), and independent construction (autonomous student text production). Research demonstrates that genre-based approaches significantly enhance rhetorical awareness, argument organization, and text quality, helping students understand explicit relationships between claims, reasons, and evidence.

➤ *Research Gap and Research Questions*

Despite documented effectiveness of genre-based pedagogy and growing AI-assisted writing research, significant gaps remain. Most AI research focuses on surface-level linguistic improvements rather than operationalized critical thinking dimensions. Genre-based pedagogy studies rarely integrate ChatGPT as a dialogic scaffolding partner. Critical thinking research typically employs conventional approaches without AI support. Furthermore, context-specific quasi-experimental studies examining AI integration in ESP academic writing within Indonesian higher education remain limited.

In East Java, 344 active higher education institutions (17 public, 327 private) serve approximately 947,251 students with teacher-to-student ratios often ranging from 1:32 to 1:40, severely limiting individual writing feedback. Digital literacy varies considerably, and only 18% of private universities hold top accreditation. These contextual factors create urgent need for scalable, reflective feedback systems.

This research addresses these gaps by: (1) developing an integrative model combining ChatGPT, genre-based pedagogy, and operationalized critical thinking dimensions; (2) positioning ChatGPT as a dialogic cognitive partner throughout the Teaching-Learning Cycle; (3) employing

mixed-methods sequential explanatory design combining quasi-experimental quantitative analysis with qualitative investigation; and (4) providing empirical evidence for AI-based pedagogy development in Indonesian higher education contexts.

II. LITERATURE REVIEW

➤ *AI as Cognitive Partner: Beyond Technical Tools*

ChatGPT's development has fundamentally shifted the role of technology in writing instruction from simple grammar checking to cognitive partnership. AI-generated feedback significantly enhances essay organization and idea coherence; however, argumentative depth requires explicit pedagogical integration for reasoning enhancement.

A critical concern is 'metacognitive laziness'—the tendency of students to reduce deep reflection when AI is used without instructional regulation. While writing performance may improve, cognitive processes may become superficial if AI functions as thought replacement. Therefore, AI must be designed as cognitive scaffolding rather than cognitive replacement. Generative AI demonstrates complex relationships with learning autonomy: it strengthens self-directed learning within reflective frameworks but risks epistemic dependency without metacognitive regulation.

Students view ChatGPT as effective for brainstorming but less helpful for argument quality evaluation. Notably, combining peer feedback with ChatGPT feedback produces more reflective revision than either approach alone, positioning AI as an immediate response system while human interaction strengthens argumentative dialogue. This evidence supports positioning ChatGPT within structured pedagogical frameworks rather than as standalone tool.

➤ *Systemic Functional Linguistics and Genre-Based Pedagogy Framework*

Systemic Functional Linguistics conceptualizes language as a social semiotic system realizing three metafunctions: ideational (representing cause-effect and logical structure), interpersonal (showing writer positioning), and textual (managing discourse cohesion and coherence). Genre-based pedagogy operationalizes this framework through a structured Teaching-Learning Cycle.

Research demonstrates that genre-based approaches significantly increase claim structure, reason elaboration, and evidence integration. Toulmin Model integration within genre scaffolding enhances students' warrant and rebuttal usage. Longitudinal studies show argumentative element mastery increases with explicit genre frameworks. Integrating genre-based approaches with critical pedagogy strengthens higher-order thinking skills.

However, despite proven effectiveness, research integrating ChatGPT within the Teaching-Learning Cycle remains limited. Most genre-based studies rely on traditional instructor-student interaction without systematic exploration

of AI as a dialogic partner. This represents a critical research gap the current study addresses.

➤ *Critical Thinking in Argumentative Academic Writing*

Critical thinking in academic writing encompasses argument evaluation, problem analysis, and decision justification. Research shows EFL students often struggle with counter-argument development and logical rebuttal construction. Brain-based instructional programs significantly enhance decision justification. AI-based collaborative inquiry increases argument evaluation through reflective technology-mediated dialogue. Motivational regulation strategies mediate improvements in writing proficiency and critical reasoning. Integrating automated writing evaluation with generative AI feedback significantly enhances higher-order thinking. These findings collectively indicate that critical thinking development requires explicit scaffolding, reflective dialogue, and systematic argumentative structure—all of which can be enhanced through structured AI integration.

➤ *ESP Academic Writing and Disciplinary Context*

ESP academic writing demands argumentative structure adaptation aligned with disciplinary conventions. Research identifies student difficulties in organization, discourse development, and evidential support. Genre-based group writing approaches significantly enhance argumentative proficiency among EFL students. ChatGPT increases both ESP writing proficiency and learner autonomy. Generative AI intervention enhances argumentative coherence within university contexts. However, AI integration specifically targeting critical thinking dimensions in ESP contexts remains systematically under-researched, representing the gap this study addresses.

III. METHODOLOGY

➤ *Research Design*

This study employs a mixed-methods sequential explanatory design, combining quantitative and qualitative approaches to investigate ChatGPT integration within genre-based pedagogy for strengthening critical thinking in ESP academic writing. The quantitative phase utilized a quasi-experimental design with pre-test and post-test measures to assess changes in critical thinking capabilities. This phase provides numerical evidence of pedagogical effectiveness. The qualitative phase employed observations, interviews, and document analysis to understand the pedagogical processes and student experiences during implementation. This sequential approach allows for comprehensive understanding: quantitative findings identify the magnitude of change while qualitative findings illuminate the mechanisms through which change occurs. This mixed-methods approach addresses both research effectiveness ("Does it work?") and mechanism questions ("How and why does it work?")

➤ *Participants and Research Sites*

This study was conducted at six private universities in East Java offering English education programs with ESP academic writing courses. Selection used purposive

sampling based on: (1) offering ESP academic writing courses, (2) possessing digital infrastructure enabling AI-based learning, and (3) instructor willingness to integrate ChatGPT. The research involved 142 students enrolled in ESP academic writing courses (61 male, 43%; 81 female, 57%) in semesters 4-5. All participants had completed foundational writing courses, ensuring basic English academic writing competency. Six instructors (one per institution) facilitated the teaching-learning cycle implementation.

For confidentiality, participants received coded identifications: institutions were coded U1-U6, and students used codes like U1-01, U2-07, etc. Instructors were coded L1-L6. This coding system was maintained throughout all analyses.

• *Ethical Approval and Informed Consent*

The source manuscript does not provide the approving ethics committee, approval number, or informed consent procedure. These details must be supplied before submission.

➤ *Data Collection Methods*

• *Pre-Test and Post-Test Essays*

Students completed argumentative essays before and after the genre-based pedagogy intervention integrated with ChatGPT. Essays were evaluated using an analytic rubric measuring three critical thinking dimensions:

- ✓ Argument Evaluation: Student ability to assess argument strength, identify logical fallacies, and recognize evidence limitations
- ✓ Problem Analysis: Student ability to identify core issues, examine multiple perspectives, and develop nuanced problem understanding
- ✓ Decision Justification: Student ability to provide evidence-based reasoning, connect claims to supporting evidence, and explain decision rationale

The rubric employed a 4-point scale (1 = emerging, 2 = developing, 3 = proficient, 4 = advanced) with detailed descriptors for each level. Two trained raters independently scored all essays, with inter-rater reliability (Cronbach's alpha = 0.87) exceeding the 0.75 threshold.

• *Classroom Observations*

Non-participant observations documented implementation of the Teaching-Learning Cycle phases (modeling, joint construction, independent construction) across 12 weeks (approximately 36 hours total observation). Observations specifically focused on:

- ✓ Pedagogical practices in implementing each Teaching-Learning Cycle phase
- ✓ Student-ChatGPT interaction patterns during writing and revision
- ✓ Student engagement in argumentative discussions
- ✓ Evidence of metacognitive reflection and critical evaluation

Detailed field notes were recorded and coded using participant identifiers (e.g., 'U2-07 used ChatGPT to evaluate claim strength during joint construction').

- *Semi-Structured Interviews*

Semi-structured interviews were conducted with 36 purposively selected students (approximately 25% of participants) representing high, medium, and low academic performers. Interviews explored student experiences using ChatGPT, perceived effectiveness, challenges, and suggestions for improvement. Interviews were audio-recorded, transcribed, and coded while maintaining confidentiality through participant identification codes.

- *Document Analysis*

Student artifacts were collected including: draft essays (initial and revised versions), ChatGPT interaction transcripts (when provided by students), and reflection notes. These documents traced writing development across the learning cycle, showing how students applied feedback and revised thinking.

➤ *Data Analysis Procedures*

- *Quantitative Analysis*

Pre-test and post-test essay scores were analyzed using paired-sample t-tests to determine statistical significance of pre-post differences. Additional analyses included:

- ✓ Effect size calculation (Cohen's d) to assess practical significance
- ✓ Descriptive statistics (means, standard deviations) for each critical thinking dimension

- ✓ Subgroup analyses examining potential differences by academic level and gender

Significance was set at $p < .05$. SPSS Version 26 was used for all statistical analyses.

- *Qualitative Analysis*

Qualitative data from observations, interviews, and documents were analyzed using thematic analysis following Braun and Clarke's six-phase procedure: (1) data familiarization through repeated reading, (2) initial coding to identify meaningful units, (3) code grouping into categories, (4) identification of overarching themes, (5) theme review and refinement, and (6) final interpretation. Analysis specifically examined how ChatGPT functioned within each Teaching-Learning Cycle phase and how it supported critical thinking development. A codebook was developed and applied systematically across data sources. To enhance credibility, triangulation across data types (observations, interviews, documents) was employed, and member checking involved instructor review of preliminary interpretations.

IV. RESULTS

➤ *Participant Characteristics*

The study involved 142 participants from six private universities in East Java. Demographic distribution was relatively balanced: 61 male students (43%) and 81 female students (57%). Participants were distributed across semesters 4 ($n=72$, 51%) and 5 ($n=70$, 49%), providing representation of intermediate academic writing students. All participants had completed foundational writing courses, ensuring comparability of baseline writing ability.

Table 1 Participant Characteristics

Characteristic	Number (n)	Percentage (%)
Male	61	43%
Female	81	57%
Semester 4	72	51%
Semester 5	70	49%

➤ *Pre-Test Results: Baseline Critical Thinking Abilities*

Pre-tests conducted before intervention implementation assessed baseline critical thinking capabilities in argumentative essay writing. Results revealed that students possessed moderate-to-low critical thinking abilities overall ($M = 2.64$, $SD = 0.49$ on a 4-point scale). Most notably, students demonstrated particular weakness in

decision justification ($M = 2.58$, $SD = 0.47$), indicating difficulty providing strong, evidence-based reasoning in argumentative writing. Argument evaluation abilities ($M = 2.71$, $SD = 0.52$) and problem analysis ($M = 2.64$, $SD = 0.49$) were slightly stronger but remained in the developing range. This baseline assessment indicated substantial room for improvement and justified the pedagogical intervention.

Table 2 Pre Test Critical Thinking Scores

Critical Thinking Dimension	Mean (M)	Standard Deviation (SD)
Argument Evaluation	2.71	0.52
Problem Analysis	2.64	0.49
Decision Justification	2.58	0.47
Overall Mean	2.64	0.49

➤ *Post-Test Results: Changes in Critical Thinking Abilities*

Following the 12-week intervention implementing ChatGPT-integrated genre-based pedagogy, post-test essays

revealed substantial improvements across all critical thinking dimensions. Overall critical thinking increased significantly ($M = 3.75$, $SD = 0.46$), representing a gain of

1.11 points on the 4-point scale. Specific dimension improvements included:

- Argument Evaluation increased from M = 2.71 to M = 3.82 (gain = 1.11 points, 41% improvement)
- Problem Analysis increased from M = 2.64 to M = 3.75 (gain = 1.11 points, 42% improvement)

- Decision Justification increased from M = 2.58 to M = 3.69 (gain = 1.11 points, 43% improvement)

These improvements shifted students from 'developing' to 'proficient' range for all dimensions, indicating meaningful progress toward independent, sophisticated argumentation.

Table 3 Posttest Critical Thinking Scores

Critical Thinking Dimension	Mean (M)	Standard Deviation (SD)
Argument Evaluation	3.82	0.48
Problem Analysis	3.75	0.46
Decision Justification	3.69	0.44
Overall Mean	3.75	0.46

➤ Statistical Significance Testing

Paired-sample t-tests were conducted to determine whether pre-to-post improvements reached statistical significance. All three critical thinking dimensions showed statistically significant improvements ($p < .001$ for all

dimensions). Effect sizes (Cohen's d) ranged from 2.13 to 2.28, representing very large practical effects. These results provide strong evidence that the ChatGPT-integrated genre-based pedagogy intervention produced meaningful improvements in critical thinking.

Table 4 Paired Sample Test Results and Effect Sizes

Dimension	Pre-Test M	Post-Test M	t-value	p-value (Cohen's d)
Argument Evaluation	2.71	3.82	12.14	<.001 (2.24)
Problem Analysis	2.64	3.75	11.68	<.001 (2.15)
Decision Justification	2.58	3.69	10.95	<.001 (2.13)

➤ Classroom Observation Findings

Observations during the teaching-learning cycle revealed important implementation patterns:

- **Modeling Phase:** Students actively analyzed argumentative essay structures and identified rhetorical strategies. When using ChatGPT to compare essay variations, students engaged in deeper discussion about how different organizational patterns affect argument persuasiveness.
- **Joint Construction Phase:** Collaboration between instructors and students intensified when ChatGPT was integrated. Students asked ChatGPT specific questions about argument strength ('Is this claim too broad?', 'Does my evidence support this claim?'), which prompted reflective discussion about reasoning quality.
- **Independent Construction Phase:** Students autonomously used ChatGPT for revision, asking questions like 'What counterarguments should I consider?' and 'Is my evidence sufficient?' This represented evidence of metacognitive engagement—students evaluating their own thinking.

Across all phases, ChatGPT functioned as an immediate feedback source, allowing students to obtain rapid responses and adjust thinking, while instructor feedback provided pedagogical context and deeper analysis.

➤ Interview Findings: Student Perceptions and Experiences

Semi-structured interviews ($n=36$, 25% of sample) revealed consistent student perceptions about ChatGPT integration. Analysis identified four major themes:

• Theme 1: ChatGPT as Idea Development Partner

Students frequently reported that ChatGPT helped when experiencing writing anxiety or initial confusion about essay development. Student U1-04 stated: 'Usually when asked to write essays I get confused about where to start. Sometimes I have ideas but don't know how to explain them. When I use ChatGPT, I ask about the topic and then I get a picture of how to develop my arguments.'

- ✓ Similarly, U2-07 noted: 'The most helpful thing is when making an outline. I usually ask ChatGPT what arguments I could use. From that I can more easily arrange paragraphs one by one.' This demonstrates that ChatGPT functioned as a cognitive scaffold for organizing initial thoughts.

• Theme 2: ChatGPT as Argument Quality Evaluator

Several students reported that ChatGPT helped evaluate argument strength. Student U3-08 explained: 'When I put my paragraphs into ChatGPT, it tells me if my arguments are strong enough. Sometimes it says my reasons aren't clear or I need additional examples.'

- ✓ Student U4-05 revealed a shift in focus: 'Previously I was only focused on correct grammar. But after using ChatGPT, I started realizing that arguments also need to be logical. Sometimes I think my arguments are good, but they can actually be improved.'

This theme indicates that ChatGPT prompted metalinguistic awareness—students began evaluating not just surface features but logical coherence.

- *Theme 3: Understanding Claims, Reasons, and Evidence Relationships*

Several students explicitly mentioned increased understanding of argumentation structure. Student U5-09 stated: 'Usually I just write my opinion without much evidence. But when I ask ChatGPT, it often suggests adding examples or data. So my writing feels stronger.'

U6-03 reflected: 'I learned that in essays it's not enough to just say "I agree" or "I disagree." There must be reasons and clear examples.' This demonstrates that ChatGPT interactions promoted understanding of argumentative structure—the explicit linking of claims, warrants, and evidence.

- *Theme 4: Metacognitive Caution About Over-Reliance*

Importantly, students also expressed metacognitive awareness about potential pitfalls. Student U4-10 cautioned: 'I think ChatGPT helps, but I still have to think for myself. Sometimes its suggestions don't fit my ideas.'

U5-12 explained their approach: 'I don't just copy answers from ChatGPT. Usually I just look at its suggestions, then I try to adjust them with my own writing.' These reflections indicate that students recognized the importance of maintaining cognitive agency and evaluating AI suggestions critically.

Student U1-09 reported increased confidence: 'Previously I felt essay writing was very difficult. But after using ChatGPT several times, I feel more confident because I know how to improve my writing.'

U6-07 noted motivational benefits: 'Usually I give up quickly if my writing is bad. But with ChatGPT I feel I have a tool that can help me improve my writing.'

These findings collectively indicate that students perceived ChatGPT as a supportive learning tool that enhanced their writing confidence and critical thinking engagement.

V. DISCUSSION

The findings of this research demonstrate that integrating ChatGPT within genre-based pedagogy produces significant improvements in critical thinking abilities among ESP academic writing students. Both quantitative and qualitative evidence support this conclusion, illuminating both the magnitude and mechanisms of change.

- *Improved Argument Evaluation: Moving Beyond Surface Features*

The 41% improvement in argument evaluation (2.71 to 3.82) demonstrates that students became substantially more capable of assessing argument strength and identifying logical weaknesses. In genre-based pedagogy terms, this reflects enhanced understanding of the ideational metafunction—how language constructs logical relationships and causal structures. ChatGPT functioned as a dialogic feedback mechanism enabling students to

iteratively test their arguments. Unlike traditional feedback which occurs after writing completion, ChatGPT provided immediate responses, allowing students to evaluate argument quality during the writing process. This immediate feedback created what researchers term "cognitive scaffolding"—supporting thinking processes without replacing them. Importantly, qualitative data revealed that students shifted from surface-level grammar focus toward deeper logical analysis. This aligns with research on AI as cognitive offloading, where technology frees cognitive resources for higher-order thinking rather than mechanical tasks. Students could delegate routine aspects to ChatGPT while concentrating on argumentative rigor.

- *Enhanced Problem Analysis: Deeper Issue Understanding*

The 42% improvement in problem analysis (2.64 to 3.75) indicates students developed more sophisticated understanding of argumentative contexts and multi-perspective thinking. Problem analysis requires identifying core issues, recognizing complexity, and developing nuanced understanding—all metacognitive skills. During the teaching-learning cycle, ChatGPT served as a thinking partner. In the joint construction phase, when students and instructors co-built arguments, ChatGPT provided alternative framings and perspectives. Student observations confirmed this: they reported asking ChatGPT "what counterarguments should I consider?" This demonstrates that ChatGPT scaffolded perspective-taking, a key critical thinking dimension. The improvement also reflects successful genre-based pedagogy operationalization. By explicitly teaching argumentative structure through modeling and joint construction, students internalized how problems should be framed and analyzed. ChatGPT reinforced these structures through consistent, immediate feedback on how well students applied genre knowledge.

- *Strengthened Decision Justification: Evidence-Based Reasoning*

The 43% improvement in decision justification (2.58 to 3.69) is particularly significant, as this dimension—providing evidence-based reasoning—proved the weakest at baseline. This improvement indicates that students developed stronger capacity for supporting claims with relevant, credible evidence. The interpersonal metafunction, according to SFL, reflects the writer's commitment to and evaluation of claims. ChatGPT interactions promoted this by prompting students to justify their positions: "Why do you believe this?" "Is your evidence sufficient?" These questions made explicit the evaluative work required in academic argumentation. Notably, some students reported pre-intervention focus on grammatical correctness without argumentative substance. Post-intervention, they demonstrated awareness that "arguments also need to be logical." This cognitive shift—from form-focused to meaning-focused writing—represents genuine critical thinking development.

➤ *ChatGPT as Scalable Feedback System in Resource-Constrained Contexts*

An important contextual contribution of this research concerns the application of AI within Indonesian higher education. In many East Java institutions, teacher-to-student ratios of 1:32 to 1:40 severely limit individual instructor feedback. This research demonstrates that ChatGPT can provide scalable feedback addressing this constraint. Unlike hiring additional instructors, deploying ChatGPT requires no additional human resources yet provides immediate, individualized feedback to all students. During independent construction phases, students could receive feedback instantly rather than waiting days for instructor comments. This represents a practical solution to persistent resource constraints in higher education. However, findings also emphasize that ChatGPT functions most effectively within structured pedagogical frameworks. Qualitative data showed students benefited most when their instructors explicitly taught how to use ChatGPT reflectively. Without instructional guidance, technology risks becoming a shortcut rather than learning tool.

➤ *Critical Awareness About Metacognitive Regulation*

An important finding concerns student recognition of potential pitfalls. Several students noted that ChatGPT suggestions required personal evaluation: "I don't just copy answers from ChatGPT... I try to adjust them with my own writing." This metacognitive caution addresses concerns in the literature about "metacognitive laziness"—the tendency to reduce reflection when technology is available. These students demonstrated maintained cognitive engagement despite AI availability. This suggests that when AI is integrated within pedagogically structured curricula emphasizing critical thinking, students develop healthy skepticism about AI outputs rather than blind acceptance. The importance of this finding cannot be overstated. It indicates that the genre-based pedagogy framework—which emphasizes explicit understanding of linguistic choices and rhetorical purposes—promoted critical evaluation of AI-generated suggestions. Students understood that ChatGPT outputs required human judgment and contextual adaptation.

➤ *Alignment with Systemic Functional Linguistics Framework*

Theoretically, improvements across all critical thinking dimensions align with SFL's three metafunctions. The ideational metafunction (representing logical structure) improved through enhanced argument evaluation. The interpersonal metafunction (showing writer stance and evaluation) improved through strengthened decision justification. The textual metafunction (organizing discourse coherence) was supported throughout through genre-based scaffolding. This theoretical alignment suggests that the integrated model operates coherently. Critical thinking development was not ad hoc but grounded in explicit linguistic-functional theory operationalized through structured pedagogy. ChatGPT, rather than contradicting this framework, reinforced it by providing consistent feedback on how well students' linguistic choices realized argumentative purposes.

VI. CONCLUSION

This research demonstrates that integrating ChatGPT within genre-based pedagogy significantly strengthens critical thinking in ESP academic writing. Quantitative analysis revealed statistically significant improvements ($p < .001$) across three critical thinking dimensions with very large effect sizes (Cohen's $d = 2.13$ - 2.28). Qualitative findings illuminated mechanisms through which improvement occurs: ChatGPT functioned as a dialogic cognitive partner facilitating idea development, argument evaluation, evidence elaboration, and reflective revision. The research makes several important contributions to the field. First, it provides empirical evidence for an integrated model combining AI technology, genre-based pedagogy, and operationalized critical thinking instruction. This addresses a significant research gap where previous studies examined components separately but not integrally. Second, the research demonstrates that ChatGPT functions most effectively within structured pedagogical frameworks emphasizing explicit critical thinking, not as a standalone tool. Third, the research contributes contextually-specific evidence for Indonesian higher education, where resource constraints and digital transformation urgently require innovative solutions.

The findings suggest that ChatGPT can serve as a scalable, cost-effective feedback mechanism addressing persistent constraints in higher education writing instruction. However, effectiveness depends critically on pedagogical design emphasizing explicit teaching of critical thinking, structured support through the teaching-learning cycle, and cultivation of metacognitive awareness about appropriate AI use. The research also revealed encouraging evidence that students develop healthy critical awareness about AI limitations. Rather than creating dependency, well-designed AI integration can promote greater metacognitive engagement and critical evaluation of information sources. This finding contradicts some concerns in the literature about AI creating passive learners; instead, students demonstrated maintained cognitive agency and careful evaluation of AI suggestions.

VII. LIMITATIONS

➤ *Despite Important Contributions, Several Limitations Warrant Acknowledgment:*

- **Limited Institutional Scope:** The study involved only six private universities in East Java, potentially limiting generalization to national contexts or public universities. Institutional characteristics, curriculum design, and digital readiness vary considerably across contexts.
- **Quasi-Experimental Design:** The absence of a control group prevents definitively attributing improvements solely to ChatGPT integration. Other factors—including student motivation, instructor quality, and prior learning experiences—may contribute to observed improvements. A true experimental design with randomized control group would strengthen causal claims.

- **Limited Assessment Scope:** Critical thinking was assessed through argumentative essay writing only. Other critical thinking manifestations (problem-solving, collaborative reasoning, decision-making in authentic contexts) were not examined. The findings thus reflect critical thinking in this specific genre rather than critical thinking comprehensively.
- **Short-Term Implementation:** The 12-week implementation period provides evidence of short-term effects. Long-term impacts on critical thinking development and potential skill retention after intervention conclusion remain unknown.
- **Potential Selection Bias:** Students willing to use new technology and instructors willing to implement innovative pedagogy may differ systematically from the broader population. Motivation and early-adopter orientation may explain some findings.
- **Measurement Considerations:** While rubric reliability was adequate (Cronbach's $\alpha = 0.87$), the rubric was not externally validated against established critical thinking measures. Future research might employ validated instruments for comparison.

ACKNOWLEDGMENTS

The authors would like to thank all students and instructors who participated in this study and contributed their time and experiences to the research. The authors also acknowledge the support of the participating institutions in facilitating the implementation of the study.

➤ *Conflict of Interest*

The authors declare that there is no conflict of interest regarding the publication of this article.

➤ *Funding*

This research received no external funding.

REFERENCES

- [1]. Abdel Latif, M. M., et al. (2024). Systemic functional linguistics in genre-based pedagogy: Enhancing argumentative writing development. *Journal of Academic Writing*, 14(3), 245–268.
- [2]. Ananda, R., Mardiah, R., & Fajaryani, N. (2024). Writing argumentative essays: Jambi EFL students' challenges and strategies. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 9(2). <https://doi.org/10.21093/ijeltal.v9i2.1722>
- [3]. Ayoub, A., et al. (2025). Academic integrity concerns in AI-assisted writing: Policy perspectives. *Higher Education Review*, 57(2), 134–157.
- [4]. Banihashem, S. K., et al. (2024). AI-generated feedback in academic writing: Impact on coherence and organization. *Computers and Education*, 198, 104–119.
- [5]. Chen, L., et al. (2025). GenAI-based collaborative inquiry for developing critical thinking in writing. *Journal of Educational Technology*, 31(4), 612–631.
- [6]. Dong, Y. (2025). Motivational regulation in AI-supported writing instruction. *Learning and Instruction*, 88, 101–115.
- [7]. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, T., Shen, Y., Li, X., & Gašević, D. (2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 13(6), 13544. <https://doi.org/10.1111/bjet.13544>
- [8]. Ghanbari, B., & Salari, M. (2022). Critical thinking in argumentative essay writing: EFL learners' strengths and weaknesses. *TESL Canada Journal*, 39(1), 56–78.
- [9]. Hao, M., et al. (2026). Integrating automated evaluation and generative AI feedback: Effects on higher-order thinking. *Studies in Educational Evaluation*, 74, 1–12.
- [10]. Huang, D., & Teng, Y. (2025). Combining peer and AI feedback for reflective revision in academic writing. *Journal of Writing Research*, 17(2), 156–181.
- [11]. Insuwan, S., & Thongrin, S. (2025). Genre-based critical pedagogy and higher-order thinking development. *Language Teaching Research Quarterly*, 48(3), 267–289.
- [12]. Juhana, J., et al. (2023). Challenges in ESP academic writing: Organization and logical progression. *ESP Across Cultures*, 20(2), 112–134.
- [13]. Karataş, F., Abedi, F. Y., Gunyel, F. O., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT's effect on foreign language learners. *Computer Assisted Language Learning*, 27(3). <https://doi.org/10.1007/s10639-024-12574-6>
- [14]. Lestari, R. (2024). Examining logico-semantic relation in Indonesian EFL students' writing in schooling genres: SFL perspective. *Indonesian Journal of English Language Education*, 11(2). <https://doi.org/10.15408/ijee.v11i2.36728>
- [15]. Li, D., & Wang, Y. (2024). Translanguaging as a mediational tool in the process of feedback in second language writing. *Contemporary Linguistics*, 6(1). <https://doi.org/10.1515/CJAL-2024-0108>
- [16]. Lim, J. M., & Go, C. (2025). Integrating ChatGPT into teacher feedback: Practical insights for L2 writing instruction. *English Journal*, 291(15). <https://doi.org/10.55593/ej.29115int>
- [17]. Liu, L., & Arumugam, N. (2025). Developing argumentative writing proficiency in Chinese EFL classrooms: A genre-based group writing approach. *Universal Journal of Social Sciences and Humanities*, 4(1), 36. <https://doi.org/10.53797/ujssh.v4i1.36.2025>
- [18]. Malik, N., Kousar, A., & Arshad, K. (2025). ChatGPT in education: Learning outcomes and facilitating knowledge acquisition. *International Journal of Social Sciences*, 4(2), 131. <https://doi.org/10.63544/ijss.v4i2.131>
- [19]. Maru, M., Moge, T., & Sabudu, D. (2025). Conceptualizing speech texts as a learning model to foster critical thinking skills in argumentative writing classes at higher education institutions. *Jurnal*

- Metalingua, 29(2).
<https://doi.org/10.52152/mdk29z20>
- [20]. Moyano, E. I. (2025). Enseñanza de la lectura, la escritura y la oralidad académicas y profesionales: Aspectos nodales para el desarrollo de un programa en el nivel superior. *Revista Digital de Políticas Lingüísticas*. <https://doi.org/10.63790/c9gf5p30>
- [21]. Nguyen, T. (2026). Genre and register in medical research abstracts: A systemic functional linguistic approach. *Journal of Functional Syntax*, 5(5), 5508. <https://doi.org/10.63023/2525-2445/jfs.ulis.5508>
- [22]. Ozdere, M. (2025). AI in academic writing: Assessing the effectiveness, grading consistency, and student perspectives of ChatGPT and You.com for EFL students. *International Journal of Teacher Education*, 1001. <https://doi.org/10.46328/ijte.1001>
- [23]. Rachmawati, D., & Fadhilawati, D. (2024). Exploring the challenges faced by ESP students in writing academic essay. *Journal of English Teaching and Learning for Education*, 4(2), 2962. <https://doi.org/10.47766/jetlee.v4i2.2962>
- [24]. Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Language and Teaching*, 6(1), 9. <https://doi.org/10.37074/jalt.2023.6.1.9>
- [25]. Street, B. (2003). What's 'new' in new literacy studies? Critical approaches to literacy in theory and practice. *Current Issues in Comparative Education*, 5(2), 11369. <https://doi.org/10.52214/cice.v5i2.11369>
- [26]. Wahyudi, A., Pamuji, A., Martuti, R., & Hapradespa, A. (2024). A need analysis of ESP for nursing science students at STIK Bina Husada Palembang. *International Journal of English Applied Language*, 4(3), 4945. <https://doi.org/10.47709/ijeal.v4i3.4945>
- [27]. Zhang, W., Singh, K. K. M., & Ibrahim, M. N. A. (2024). The analysis of schematic structure in students' argumentative essays. *Malaysian Journal of Social Sciences and Humanities*, 9(7), 2869. <https://doi.org/10.47405/mjssh.v9i7.2869>
- [28]. Zhai, X., Chu, X., Chai, C. S., Jong, M. S.-Y., Starčić, A. I., Spector, M., Liu, J., Jing, Y., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Journal of Educational Computing Research*, 2021, 8812542. <https://doi.org/10.1155/2021/8812542>
- [29]. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39. <https://doi.org/10.1186/s41239-019-0171-0>