RESPONSE TO EDITORS AND REVIEWERS (ROUND 2)

Manuscript Title:
Traditional Writing and E-Writing Using Gen-AI to Empower Students' Critical Thinking in Argumentative Text: A Case Study in Indonesian EFL Higher Education

Journal:
English Review: Journal of English Education (ERJEE)

Manuscript ID: [235]

Authors:
Dedeh Rohayati, Etika Rachmawati, R. Bunga Febriani, and Bambang Ruby Sugiarto
Department of English Education, Faculty of Teacher Training and Education Universitas Galuh, Ciamis, Indonesia

Dear Editor and Reviewers,

We sincerely thank the Editor and the Reviewers for their careful reading of our manuscript and for providing constructive and insightful comments. We greatly appreciate the time and effort invested in reviewing our work. The comments have helped us identify important areas for improvement and have enabled us to strengthen the manuscript substantially.

We have carefully considered all comments and suggestions and have revised the manuscript accordingly. In particular, we have revised Language Quality, Framing, Literature, References, and Scholarly Presentation (Reviewer 1) and Methodology, Data Analysis, Trustworthiness, Ethics, and Interpretation of Findings (Reviewer 2).

Following the reviewers' suggestions, we have provided a detailed point-by-point response below. For ease of reference, the reviewers' comments are presented in **bold**, followed by our responses.

We hope that the revisions have addressed the reviewers' concerns and have substantially strengthened the manuscript.

REVIEWER A — Focus: Language Quality, Framing, Literature, References, and Scholarly Presentation

Overall assessment. The authors have clearly invested substantial effort in expanding the reference list, reframing the research gap, and restructuring the paper. The topic is timely and the Indonesian EFL higher education context is a genuine contribution. However, this second round reveals a persistent gap between what the response letter claims to have done and what the manuscript actually shows. The revision appears to have been assembled hastily: language problems remain throughout, several sections are duplicated, and a

number of "Revision Made" statements are not verifiable in the text. The manuscript is not yet at a publishable standard of scholarly presentation.

Language and academic style. This remains the most immediately visible weakness. Despite the response letter's claim of "comprehensive language and style editing," the manuscript still contains basic grammatical and lexical errors: "the existential of critical thinking is highly required"; "it makes students stuck in critical thinking"; "exemplipied"; "demonstarte crititcal thinking"; "This imply significanly"; "the present study weakened by the category of questions"; "students' awareness of 1) traditional writing..., 2) e-writing leading to..., 3) Gen-AI use in polishing..." (a run-on list inside a sentence); "The 24 third year students enrolling Academic Writing Class in English Education Program. were selected purposively." (a sentence fragment with a stray period). Terminology is also unstable: "Gen-AI," "Gen AI," "GEN-AI," and "Gen Ai" all appear; "e-writing" and "electronic-based writing" are used interchangeably; "argumentative text," "argumentative essay," and "argumentative writing" shift without definitional clarity. The claim of a completed language edit is therefore not supported by the evidence, and I strongly recommend a certificate of professional editing be provided with the next submission.

Authors' response:

a. Existential -→ existence
b. it makes students stuck in critical thinking -→students' development of critical thinking are impeded (Hassen, 2025) due to the fact that students are too much to rely on Gen-AI...
c. exemplipied→exemplified
d. demonstarte crititcal→demonstrated critical
e. significanly→ significantly
f. students' awareness of →, i.e., students aware of…
g. "The 24 third year students enrolling Academic Writing Class in English Education Program were selected purposively → Period between Program and were has been deleted.
h. Terminology on Gen-AI has been stabilized.
i. Keyword e-writing is used in abstract, the electronic-based writing is omitted.
j. Argumentative text is employed consistently to meet definitional clarity.

Title and abstract. The revised title now appropriately signals the Indonesian EFL higher education context, and this is an improvement. The abstract, however, still does not tell the reader which Gen-AI tools were used, over what period, or with what prompts. The final recommendation ("interview questions category of feeling, contrast and construct forming") will be opaque to most readers unless the terms are briefly defined; as written, it reads as specialized jargon rather than a research direction. The abstract also contains sentence fragments and inconsistent tense.

Authors' response:

a. The differences among feeling, contrast and construct forming (Cohen et al., 2013) can be elaborated as follow:

- Feeling: to dig out the emotional experiences

How did you feel when using AI during the argumentative writing process?

- Contrast: to dig out the different experiences

How was your experience different when writing with AI compared with writing without AI?

- Construct-forming; to dig out the criteria/ meaning/ concept

In your view, what makes the use of AI helpful or unhelpful for developing your argumentative writing?

So, the three categories are considered as the way to construct the interview questions in order to obtain the richer data.

b. The tense of abstract has been revised as suggested by Jiang & Hyland (2023) in that to meet the communicative function: Present tense is used for what the study/article is about, established knowledge, current claims, interpretations, and implications; past tense is used primarily for completed research procedures and events. Results may appropriately use either past or present depending on whether the writer emphasizes the completed finding or the current claim conveyed by the finding. Finally, present tense is used in this abstract to state the recommendation.

Literature review. The expanded review is broader, but it remains largely enumerative rather than synthetic. Paragraphs proceed by listing who said what (Galbraith & Baaijen, Naveed et al., Gerald & Joseph, Anderson et al., Fonkamo & Zeru, Nawi et al., Axrorova, Zheng et al.) without building a cumulative argument that leads to the study's rationale. More seriously, several citations appear only loosely relevant or are used decoratively: Yamini et al. (2025) on server log management, Li et al. (2024) on hyperledger technology in FinTech, Song & Kim (2024) on open innovation in high-tech firms, Chi et al. (2021) on clinician review of patient records, and—most problematically—Oppenheim (2018), a treatise on international law, cited to support "prompting." This last case appears to be a citation error rather than a substantive source. Sources that do not directly inform the argument should be removed, and any source used to support a claim about prompt engineering should be replaced with relevant work already in the list (Walter, 2024; Korzynski et al., 2023; Yin & Jiahao, 2024).

Authors' response:

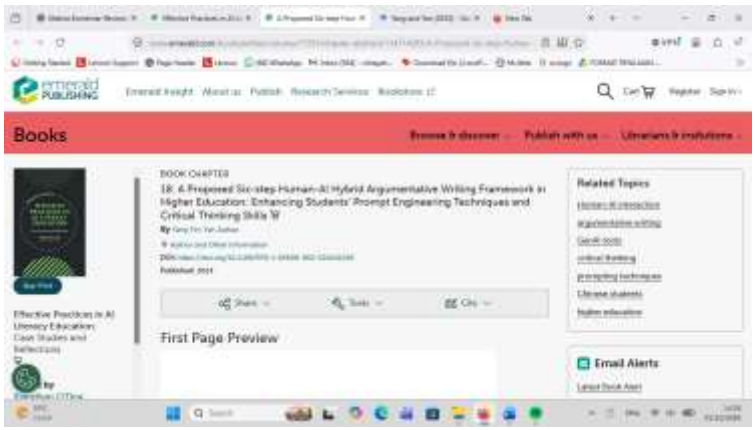
a. Paragraphs proceed by listing have been revised in INTRODUCTION, paragraph 1 and 2.
b. Supporting claim about prompt engineering has been completed INTRODUCTION, at the end of paragraph 6.

Research gap and novelty. The reviewer's request to soften the categorical gap claim has not been honored. The manuscript still states that "non of them reported the combination of traditional writing and e-writing using Gen-AI in composing argumentative text to optimize critical thinking using the model of four stages." This is exactly the over-claim flagged in Round 1. Additionally, the

reviewer specifically asked the authors to engage Yang and Yan (2025). No such reference appears in the reference list; instead, Yin and Jiahao (2024) appears, which was already cited previously. The response letter's claim to have incorporated Yang and Yan (2025) is therefore unverifiable and should be corrected or the source properly added and discussed.

Authors' response:

a. The categorical gap claim "…non of them reported…" has been soften into "…very limited study reported..." in INTRODUCTION, paragraph 11, line 10.
b. This image represents reviewer's request on Round 1(Yang and Yan (2025) proposed a six-step human-AI hybrid argumentative writing framework that includes prompt skills for ensuring synergistic interaction with Gen-AI and critical thinking skills for critically evaluating ChatGPT responses), so the author's cited Yin, Y., & Jiahao, Y. (2024) with similar research.



Reference list accuracy and APA 7. Several concrete problems require correction. (1) Reichenbach (2001a) and Reichenbach (2001b) are listed as two separate entries with identical titles and no publisher or source—these are duplicate entries for one work and should be merged and completed. (2) Bloom (1956) and Anderson and Krathwohl (2001) are cited in the text but absent from the reference list. (3) Yin (2018) is incomplete ("Sixth edition (Six Edition)" with no publisher). (4) Marrone et al. (2025) carries the same DOI as Margetis et al. (2021), indicating a copy-paste error. (5) "Valerio, B., & Theresa, M. (2025)" appears to have surname and given name reversed. (6) Nadeem et al. (2025) lacks a source or DOI. (7) In-text citations contain mechanical errors: "(Darmawansah, et al., 2025(" with an extraneous comma and unclosed parenthesis; "(Reichenbach, 2001)" in one place and "(Reichenbach, 2001a)" in another; inconsistent use of "Yin and Jiahao (2024)" versus "(Yin & Jiahao, 2024)." (8) Sources dated 2026 should be verified as formally published, with volume, issue, and DOI confirmed. Finally, the requirement of at least 80% of references from 2023–2026 should be documented; my own count suggests the actual proportion is closer to 70–75%, given retained sources such as Brown (2000), Chaffee and Carlson (2014), Cohen et al. (2013), Krathwohl (2002), Reichenbach (2001), Oppenheim (2018), Ekmekçi (2018), Galbraith and Baaijen

(2018), Limpo and Graham (2020), Kim and Belcher (2020), Tanveer et al. (2020), and numerous 2021–2022 items. The authors should provide an explicit count and, if necessary, replace older sources where the concept does not require them.

Authors' response:

a. The inconsistent use of "Yin and Jiahao (2024)" versus "(Yin & Jiahao, 2024)" due to the fact that **Citing Non-Standard Author Categories** with **A Work by Two Authors is as follow:** Name both authors in the signal phrase or in parentheses each time you cite the work. Use the word "and" between the authors' names within the text and use the ampersand in parentheses, as in …Research by Wegener and Petty (1994) supports...and as in (Wegener & Petty, 1994) (Source: https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/in_text_citations_author_authors.html)
b. Bo & Le (2024), Cui (2026), Özgür Küfi (2023), and Xue et al. (2024) have replaced Brown (2000) to meet the requirement of at least 80% of references from 2023–2026. (See subsection **The sequence of traditional and e-writing argumentative essay assisted by Gen-AI to empower students' critical thinking**, paragraph 2, line 7).
c. Brown (2000) and Chaffee and Carlson (2014) have been omitted and replaced.
d. The five years latest reference (2022–2026) consists of:

2026 = 5; 2025 = 37; 2024 = 30; 2023 = 12; 2022 = 5; with the total = 89 references (81%). So, of 110 references, there are 81% references between 2022–2026. This means that references are relatively dominated by the latest sources. The most dominant reference is 2025 which covers 37 (34%).

e. Overall, the in-text citation and references haven been fixed.

Presentation and internal consistency. There are structural errors that should never have survived revision. Figure 1 is used twice: first as "Model of four stages" and later as "Students' prompt e-writing assisted by AI." The "model of four stages" is then described in the text as a three-step procedure, while the Results section refers to "five variables." Table 1, which is central to the document analysis, is placed under the heading about students' perceptions rather than in the section analysing the documents. The heading "How do they perceive towards the implemented model?" appears twice, and the paragraph beginning "It is very fortunate that students' argumentative text was more powerful after being assisted by Gen-AI…" is duplicated verbatim. These are not stylistic quibbles; they signal to the reader that the revision was not carefully proofread. All figures and tables must be renumbered, correctly placed, and referred to consistently.

Authors' response:

a. The Figure 1 and Figure 2 have been named properly.
b. The model of four stages has been completed and displayed in paragraph 5 of subsection result "**The sequence of traditional and e-writing argumentative text assisted by Gen-AI to empower students' critical thinking".**

c. Table 1 has been replaced under the section of document analysis.
d. The subsection of "How do they perceive towards the implemented model?" has been removed as one subsection with similar name od subsection.
e. To avoid duplicated verbatim, one of expression “it is very fortunate …” has been replaced by “It is notably...”

Recommendation. Major revision. The substantive direction is sound, but the manuscript cannot proceed until the language is professionally edited, the reference list is corrected and completed, the duplicated and misplaced material is removed, and the authors' response letter accurately reflects what was actually changed. If a third round reveals the same discrepancies between claims and text, rejection would be warranted.

REVIEWER B — Focus: Methodology, Data Analysis, Trustworthiness, Ethics, and Interpretation of Findings

Overall assessment. The authors have responded to the methodological comments in principle, and the addition of trustworthiness language and a limitations statement is welcome. In practice, however, the Method section remains under-developed, the reporting of results contains internal contradictions, and the central claim—that Gen-AI-assisted e-writing "empowered" students' critical thinking—is not yet adequately supported by the design or the data. Several of the revisions promised in the response letter are simply not present in the manuscript.

Sampling and participant reporting. The justification for purposive sampling is circular: participants were selected "because 1) it was based on researchers' own judgment… and 2) due to the fact that the current research involved small numbers of participants." The second reason describes the consequence of the sampling choice, not its rationale. More importantly, the numbers are inconsistent. The Method states that twenty-four third-year students were involved, while the abstract, Method, and Results indicate that nine documents and nine interviews constitute the primary data. The rationale for reducing 24 to 9 is said to rest on GPA categories (low, mid, high), yet Table 1 presents "Essay 1" through "Essay 9" without any indication of which essay belongs to which achievement level. As a result, the manuscript's repeated claims about differences "among achievers" cannot be verified by the reader. Either label the essays by achievement category (with anonymized codes) or remove the achiever-based interpretation.

Authors’ response:

a. The justification for purposive sampling is not circular any more, and the number of participants have been polished as can be seen in METHOD.
b. Content of Table 1 has been recreate as an indication of which essay belongs to which achievement level.

Ethics. There is no statement of ethics approval, informed consent, permission to analyze

student coursework, or institutional authorization. Given that the study uses students' written documents and interviews collected within a graded Academic Writing course, the omission is significant. A statement of ethical clearance and consent procedures must be added. Relatedly, the study concerns the use of Gen-AI in assessed coursework; the manuscript should clarify whether students' AI-assisted texts were used for grading, what instructions they received about permissible AI use, and how authorship and integrity were handled. The journal's required Competing Interests and Rights statements also appear to be missing from the submission package and should be completed.

Authors' response:

The ethical consideration is available in the METHOD, and competing interest before REFERENCES

Data collection. The observation procedure remains opaque. The Method says observation "was recorded," but does not state the number of sessions, their duration, the observer's role (participant or non-participant), or the form of recording (field notes, audio, video, screen capture). The reviewer's Round 1 request for this detail has not been met. Equally important, the design compares two unequal conditions. In the traditional condition, students wrote in 60 minutes with no resources; in the e-writing condition, they had access to Gen-AI, external sources, and no stated time limit. Any observed improvement in argumentative elements is therefore confounded by time, resource access, and tool availability. The manuscript must acknowledge this explicitly, and the claim should be reframed from "Gen-AI empowered critical thinking" to something more defensible, such as "Gen-AI-supported revision was associated with more complete argumentative structure." No logs of student–AI interaction were collected, so it is not possible to determine whether the added arguments, evidence, and citations originated with the students, with the AI, or with the students' selective use of AI output. This is the single most consequential methodological limitation of the study and it is not addressed anywhere in the manuscript.

Authors' response:

The revision has been made as can be seen in METHOD.

Data analysis. The three data sources are said to be analyzed separately, but the observation analysis is still described only in relation to the interview analysis, and the document analysis is described in a single sentence. For the document analysis, the coding scheme operationalizes "critical thinking" as the presence or absence of thesis statement, argument, reason, fact/data, and opinion. This is a measure of argumentative structure, not of critical thinking, and it conflates the two. If the study intends to assess critical thinking, the authors should either adopt a recognized critical thinking framework (the Paul–Elder standards of clarity, accuracy, precision, relevance, depth, breadth, logic, significance, and fairness are mentioned by one student in the interviews but are not applied in the analysis) or explicitly redefine the construct as "argumentative completeness" and adjust the title,

abstract, and conclusions accordingly. Rubric anchors, scoring criteria, and the threshold for judging an element "present" must also be provided.

Authors' response:

Thank you for this important and critical review. We agree that critical thinking and argumentative structure are conceptually distinct constructs. However, our study does not conceptualize the presence of argumentative elements as equivalent to critical thinking. Rather, we conceptualize argumentative writing as a context in which critical thinking can be reflected and examined through the quality of reasoning and the use of supporting evidence in written arguments.

Our conceptualization of critical thinking is grounded primarily Reichenbach (2001), particularly in relation to critical thinking as a form of reflective and reasoned thinking that can be demonstrated through written discourse. In this study, we focus specifically on critical thinking manifested in argumentative writing. Argumentative texts were selected because their communicative and rhetorical purposes require writers to formulate claims, provide reasons, evaluate or use evidence, and establish logical relationships among ideas. Thus, argumentative writing provides an appropriate textual context for examining how students externalize their thinking through written arguments.

Accordingly, the argumentative elements identified in our document analysis—particularly argument/claim, reason, fact/data/evidence, and their relationships—are not intended to constitute a comprehensive definition of critical thinking. Instead, they represent observable textual manifestations through which critical thinking can be examined in students' argumentative writing. In our framework, a strong and effective argument is characterized by a clear position demonstrated in thesis statement and supported by relevant reasons and appropriate factual or evidential support. These features provide observable evidence of the writer's reasoning and reflective thinking in the written text.

We acknowledge, however, that the original manuscript did not make this conceptual relationship sufficiently explicit and may have given the impression that the presence or absence of argumentative components was itself a direct measure of critical thinking. We have therefore revised the manuscript to clarify the distinction between the **construct of critical thinking** and its **observable manifestations in argumentative writing**.

We have also revised the description of the document analysis to clarify that the coding scheme was developed to identify textual evidence of critical thinking as manifested through argumentative reasoning, rather than to claim that the five elements constitute the entirety of critical thinking. In addition, we have expanded the operational definitions and coding criteria for each element and clarified the decision rules used to determine whether an element was considered present.

Regarding the Paul–Elder framework mentioned in the interview data, we appreciate the reviewer's suggestion. We acknowledge that the Paul–Elder intellectual standards provide a recognized framework for describing dimensions of critical thinking, including clarity, accuracy,

precision, relevance, depth, breadth, logic, significance, and fairness. However, the present study was theoretically developed from the conceptualization of critical thinking in Reichenbach (2001), with a specific focus on how critical thinking is manifested in written argumentative discourse. Therefore, the mention of Paul–Elder by one student in the interview data was not intended to represent the analytical framework adopted in this study. We have clarified this distinction in the revised manuscript to avoid suggesting that the Paul–Elder framework was used as an analytical framework.

Most importantly, our four-step model is intended to develop students' critical thinking through the process of producing argumentative texts. The model does not equate critical thinking with the mechanical presence of argumentative components. Rather, it guides students toward developing stronger and more effective arguments by engaging them in the use of claims, reasons, facts/data/evidence, and reflective reasoning. We have revised the manuscript to make this theoretical and pedagogical relationship more explicit.

Trustworthiness. The claim that documents were "rated by three raters resulting the trustworthiness criteria" is insufficient. Inter-rater reliability statistics (e.g., Fleiss' or Cohen's kappa, or percentage agreement) must be reported, together with information on rater training and independence. Member checking is claimed in the response letter but is not described in the manuscript; peer debriefing is asserted but not documented. The reviewer specifically requested explicit treatment of credibility, transferability, dependability, and confirmability. Those terms do not appear in the manuscript, and no dedicated trustworthiness subsection is visible. The response letter states that such a subsection was added at specific page and line locations; those locations should be verified, because they are not evident in the current text.

Authors' response:

The revision has been made as can be seen in METHOD.

Theoretical framing. The response letter promises explicit engagement with Bloom's Taxonomy, Constructivist Learning Theory, and Cognitive Load Theory. Bloom's Taxonomy appears briefly, but neither constructivist learning theory nor cognitive load theory is discussed anywhere in the manuscript. This is a clear instance in which the response letter overstates the revision. Either the theoretical discussion must be genuinely added to the Results and Discussion, or the claim in the response letter must be withdrawn.

Authors' response:

It has been revised as can be seen at the end of paragraph 2 and at the paragraph 7 subsection: How do they perceive towards the implemented model?

Reporting of results and internal inconsistencies. Several problems undermine confidence in the analysis. First, the text states that "students' documents of traditional writing argumentative were mostly did not demonstrate thesis statement (all low achievers and 1 mid achiever)," but Table 1 shows four essays with a thesis statement and five without—and, as noted, the table does not identify achievement levels. Second, the claim that "3 of 9 students from each achiever could not present fact and data" is ambiguous; the table shows three essays missing facts/data (Essays 1, 5, and 8), not three per achiever level. Third, after Gen-AI assistance, all nine essays display all five elements. This uniform pattern is suspicious and, in the absence of interaction logs, most plausibly reflects AI-completed texts rather than independently developed student reasoning. The discussion should confront this alternative explanation directly rather than presenting the pattern as self-evidently evidence of critical thinking. Fourth, the interview dates are inconsistent: quotations in the first block are labeled "1st Interview Transcription of May 13th 2026," while the text subsequently identifies the first interview as occurring on May 6th. Fifth, quotations lack participant identifiers, and it is not stated whether the interviews were conducted in Indonesian and translated, nor how translation accuracy was ensured. Sixth, "Ge-Z" appears to be a typographical error for "Gen-Z." Finally, the paragraph beginning "It is very fortunate that students' argumentative text was more powerful…" is duplicated, and the heading "How do they perceive towards the implemented model?" appears twice—errors that make the Results section difficult to follow and suggest a lack of final proofreading.

Author's Response:

This part has been fixed.

Conclusion, limitations, and implications. The revised conclusion acknowledges only one limitation: the interview question categories. The reviewer's specific requests—small interview sample, single-institution setting, reliance on self-report, and absence of a direct critical thinking assessment—are not addressed. Also missing are acknowledgments of the confounded comparison between the two writing conditions and the possibility that AI rather than students produced the improvements observed. The claimed contribution remains vague, and the pedagogical implications, while more numerous, remain generic ("teachers should not simply replace traditional writing with AI-assisted writing"). Concrete recommendations—for example, a staged hybrid writing protocol with specified prompt-training activities, source-verification checkpoints, and assessment criteria that separate argumentative quality from AI-polishing quality—would substantially strengthen this section. Given the study's focus, an AI-use declaration and a data availability statement should also be included.

Authors' Response: The authors did not use AI because it is only used as a tool searching sources. Reviewer's request has been delivered in manuscript.

Recommendation. Major revision. The manuscript must resolve the numerical and structural inconsistencies, add ethics and trustworthiness details, report inter-rater reliability, reframe the central claim to match what the design can support, and either add or withdraw the promised theoretical and limitations content. In its current form, the

study's conclusions outrun its evidence, and the discrepancies between the response letter and the manuscript raise concerns that should be resolved before any further editorial consideration.

EDITORIAL SUMMARY (for the authors)
Both reviewers recommend major revision. The recurring concerns across both reports are: (a) a substantial gap between the revisions claimed in the response letter and the changes actually present in the manuscript; (b) internal inconsistencies in numbers, figure and table numbering, headings, and duplicated text; (c) unresolved language problems; (d) reference list errors, including duplicates, missing entries, and at least one apparently incorrect source; (e) inadequate treatment of ethics and trustworthiness; and (f) a central claim about critical thinking that is not supported by the design as currently described.