



The Impact of HAINC-Integrated NotebookLM AI Feedback on EFL Students' Writing Performance at University

Winy Junianty Imradiyah¹, Nurul Hasanah², Zulfah Fakhruddin³

^{1,2,3}English Education Department, Tarbiyah Faculty,
Institut Agama Islam Negeri Parepare

Article Info

Received: 2026-05-20

Revised: 2026-06-02

Accepted: 2026-06-10

Keywords:

AI-based feedback;

EFL writing;

HAINC protocol;

NotebookLM;

technology acceptance

DOI:

10.24256/ideas.v14i1.10716

Corresponding Author:

Winy Junianty Imradiyah

winyjunianty1@gmail.com

English Education Department,
Tarbiyah Faculty, Institut Agama
Islam Negeri Parepare

Abstract

This study investigated the impact of HAINC-integrated NotebookLM AI feedback on EFL students' writing performance at IAIN Parepare. The HAINC protocol is a structured framework that facilitates active learner-AI collaborative negotiation, while the Three-Try Rule is a progressive prompting strategy that requires three self-correction attempts based on AI cues before explicit answers are provided. Unlike conventional automated writing tools that merely provide direct error corrections, NotebookLM operates uniquely as a retrieval-augmented Socratic tutor that guides independent text revisions through inquiry-based dialogue. This innovative integration addresses a critical research gap in Indonesian higher education by shifting generative AI from a passive proofreader into an autonomous cognitive learning partner, highlighting the original contribution of this pedagogical model. A quantitative pre-experimental, one-group pre-test post-test design was employed. Given the specific context of this exploratory research, a small yet purposively targeted sample of 23 English Education students participated in the study. The instruments included a paragraph writing test and a TAM questionnaire. The findings revealed a statistically significant improvement in writing scores ($p=0.003 < 0.05$), with high technology acceptance across all TAM dimensions. While this small sample size warrants a cautious interpretation of generalizability, the study concludes that HAINC-integrated NotebookLM feedback successfully enhances EFL writing performance.

1. Introduction

Artificial Intelligence (AI) has increasingly transformed educational practices, particularly in language learning and writing instruction. In English as a Foreign Language (EFL) context, writing is widely regarded as one of the most challenging language skills because students must simultaneously manage idea development, organization, vocabulary use, grammar, and mechanics. Consequently, effective feedback plays an essential role in helping learners improve their writing performance.

Traditionally, feedback in writing classrooms has been provided manually by teachers; however, this process is often time-consuming and limited by classroom conditions, especially in large classes (Alnemrat et al., 2025). Delayed feedback may reduce students' opportunities to revise their writing effectively and negatively affect their learning progress. Amidst this rapid technological adoption, serious ethical concerns regarding AI dependency and academic integrity have emerged. When students rely excessively on automated tools to generate or fix text without intellectual engagement, they risk losing their critical thinking skills and ownership of their own writing. Therefore, integrating educational technology requires a balanced pedagogical approach that secures academic honesty while maximizing learning gains.

The issue of limited writing proficiency remains a common challenge in Indonesian higher education, including at IAIN Parepare. Based on preliminary observations, many students in the English Education Study Program still experience difficulties in developing ideas, organizing paragraphs coherently, and applying accurate grammar and vocabulary in academic writing. Recent advancements in AI technology have introduced alternative approaches to writing instruction through AI-based feedback systems. Initial studies by Alghannam (2025) demonstrate that AI feedback can improve students' writing quality by providing immediate, personalized, and continuous guidance during the revision process.

In broader EFL contexts, these digital tools assist students in grammar correction and sentence organization while simultaneously encouraging learner autonomy (Bahari, 2026). Furthermore, automated systems support students' revision behavior more efficiently than delayed conventional feedback, leading to measurable learning gains (Zhu et al., 2020)

In the specific context of Indonesian higher education, EFL learners face unique cognitive and psychological hurdles when learning academic writing. Standard classroom environments often lack the space for personalized dialogue, leaving students with direct, static corrections that they tend to accept passively without deep understanding. To foster genuine learner autonomy, Indonesian students critically require a pedagogical space for active negotiation (J Eppendi, 2025).

Instead of acting as passive recipients of automated red pen corrections, they need a dynamic interaction where they can question, evaluate, and challenge feedback, thereby internalizing writing conventions more effectively.

Among recent AI innovations, Google NotebookLM has attracted considerable attention as an educational tool capable of functioning as a Socratic tutor. Unlike conventional AI proofreading systems that directly rewrite students' sentences, NotebookLM can guide learners through inquiry-based interaction and progressive questioning (Tufino, 2025). This Socratic approach directly aligns with Self-Regulated Learning (SRL) theory, which emphasizes learners' active participation, reflection, and autonomy during the learning process (Al-Kadi, 2025). By prompting students to find answers rather than supplying them instantly, the tool optimizes cognitive processing.

This scaffolding mechanism supports the core tenets of Cognitive Load Theory, as it reduces the student's immediate cognitive burden by breaking down complex revisions into manageable stages (Ahmed & Uddin, 2025). To operationalize this balance between SRL and cognitive management, the Human-AI Interactive Negotiation Competence (HAINC) protocol serves as the ideal framework. HAINC practicalizes these theories by establishing active negotiation between students and AI systems, pushing learners to critically evaluate and negotiate AI suggestions through strategies like progressive prompting and the "Three-Try Rule" (Chen, 2025). Consequently, the integration of HAINC into NotebookLM transforms theoretical autonomy into practical classroom reflection.

Previous studies have investigated the effectiveness of AI-assisted feedback in writing classrooms. Burner et al. (2025) found that AI technologies could support formative feedback practices and improve students' engagement in writing activities. Similarly, Soori et al. (2025) reported that AI feedback positively contributed to EFL learners' writing skills when combined with teacher guidance. In another study, Fleckenstein et al. (2023) conducted a meta-analysis and concluded that automated feedback significantly improved students' writing performance across different educational levels. However, most previous studies mainly focused on AI as an automated correction tool rather than as a Socratic learning partner that promotes negotiation and learner autonomy.

Furthermore, research specifically examining NotebookLM in Indonesian higher education contexts remains extremely limited. Since NotebookLM only recently became accessible through mobile platforms in 2025, empirical investigations concerning its implementation in EFL classrooms are still scarce (Santos et al., 2025). Existing studies have not sufficiently explored how HAINC-integrated NotebookLM feedback may influence students' writing performance and technology acceptance in Indonesian universities. Therefore, a research gap exists regarding the implementation of negotiated AI feedback systems in EFL writing instruction.

Based on this gap, the present study aims to investigate the impact of HAINC-Integrated NotebookLM AI feedback on EFL students' writing performance at IAIN Parepare. Specifically, the study examines whether the implementation of NotebookLM as a Socratic AI tutor significantly improves students' writing skills in terms of content, organization, language use, and grammar. In addition, the study explores students' technology acceptance toward the AI-based feedback system using the Technology Acceptance Model (TAM), including perceived usefulness, perceived ease of use, attitude toward using AI, and behavioral intention to use the technology in the future.

This study is expected to contribute both theoretically and practically to the growing discussion of AI-assisted language learning. Theoretically, the findings may enrich discussions related to Self-Regulated Learning and Cognitive Load Theory within AI-supported writing instruction. Furthermore, in line with the findings of Pertiwi et al. (2025), This research is expected to broaden the theoretical understanding of affective factors by demonstrating how AI-driven Socratic feedback can significantly lower foreign language anxiety (FLA).

Practically, the study may provide insights for lecturers and educational institutions regarding the integration of AI-based Socratic feedback systems in EFL writing classrooms. Moreover, this study contributes to the growing discussion by positioning NotebookLM not merely as an automated correction tool but as a collaborative learning partner that encourages active negotiation, critical thinking, and learner autonomy in writing development.

2. Method

This study employed a quantitative approach using a pre-experimental one-group pre-test and post-test design to investigate the impact of HAINC-Integrated NotebookLM AI feedback on EFL students' writing performance at IAIN Parepare. Given the absence of a control group, this pre-experimental setup is objectively acknowledged as an exploratory investigation; consequently, causal interpretations regarding the findings are cautiously limited, as internal validity threats such as history or maturation effects cannot be entirely isolated. This design was selected because it allows researchers to measure students' performance before and after the implementation of the treatment within a single group (Johnson & Christensen, 2024).

The participants of this study consisted of 23 students from the English Education Study Program at IAIN Parepare selected through purposive sampling. The study focused on students who experienced difficulties in several writing components, including idea development, organization, grammar, and language use. Administrative and ethical clearances for this study were formally granted through official research permits issued by the English Education Department at IAIN Parepare and the Department of Investment and One-Stop Integrated Services (Dinas Penanaman Modal dan Pelayanan Terpadu Satu Pintu) of the Parepare City Government. To ensure the protection of human subjects, strict data anonymity

was enforced throughout the research process. All student identities, academic records, and questionnaire responses were completely anonymized during data analysis, and participation in the evaluation activities remained strictly voluntary.

Two instruments were used to collect the data: a writing test and a questionnaire. The writing test consisted of pre-test and post-test tasks designed to assess students' writing performance based on four components adapted from Hyland (2019): content and ideas, organization and structure, language and style, and grammar and punctuation. To ensure high psychometric validity and scoring reliability, the writing assessment utilized a comprehensive analytical rubric derived from (Hyland, 2019), allocating balanced weightings to each of the evaluated domains.

Furthermore, an inter-rater reliability procedure was implemented, wherein the writing scripts were independently evaluated by both the researcher and an experienced EFL writing lecturer using the standardized rubric. Any minor scoring discrepancies were subsequently resolved through professional consensus. Meanwhile, the questionnaire was developed based on the Technology Acceptance Model (TAM) proposed by Davis & Granić (2024), covering perceived usefulness, perceived ease of use, attitude toward using AI, and behavioral intention to use the technology.

The questionnaire items were tested for validity and reliability before the analysis. The validity test showed that all items were valid because each correlation value exceeded the r -table value (0.413), while the reliability test indicated a Cronbach's Alpha coefficient of 0.836, demonstrating that the instrument was highly reliable.

The study was conducted over four weeks. While this four-week timeline represents a recognized structural limitation for tracking long-term language acquisition, it is justified by the rigorous intensity and high density of the treatment provided within this concentrated window. During the first week, students completed a pre-test by writing an expository paragraph. In the second and third weeks, students participated in the treatment using NotebookLM integrated with the HAINC protocol, where the AI functioned as a Socratic tutor through progressive questioning and guided revision.

In terms of classroom dynamics, NotebookLM operated strictly as the primary digital scaffolding mechanism to facilitate students' independent self-correction (Ruby et al., 2026). This AI-driven interaction did not supplant human pedagogy; rather, general supervision and instructional guidance were fully co-maintained by both the researcher and the course lecturer. This collaborative presence ensured that the researcher could oversee the technical execution of the HAINC protocol while the lecturer addressed complex linguistic queries and managed the overall classroom environment. Students revised their writing independently using the "Three-Try Rule" to encourage critical thinking and learner autonomy (Chen, 2025).

In the final week, students completed a post-test independently without AI assistance and answered the TAM questionnaire. The collected data were analyzed quantitatively using descriptive statistics and a paired sample t-test to determine whether there was a statistically significant difference between the pre-test and post-test scores.

To provide a clear visual representation of the experimental progression, the systematic operational steps executed across the four-week intervention are delineated in Figure 1:

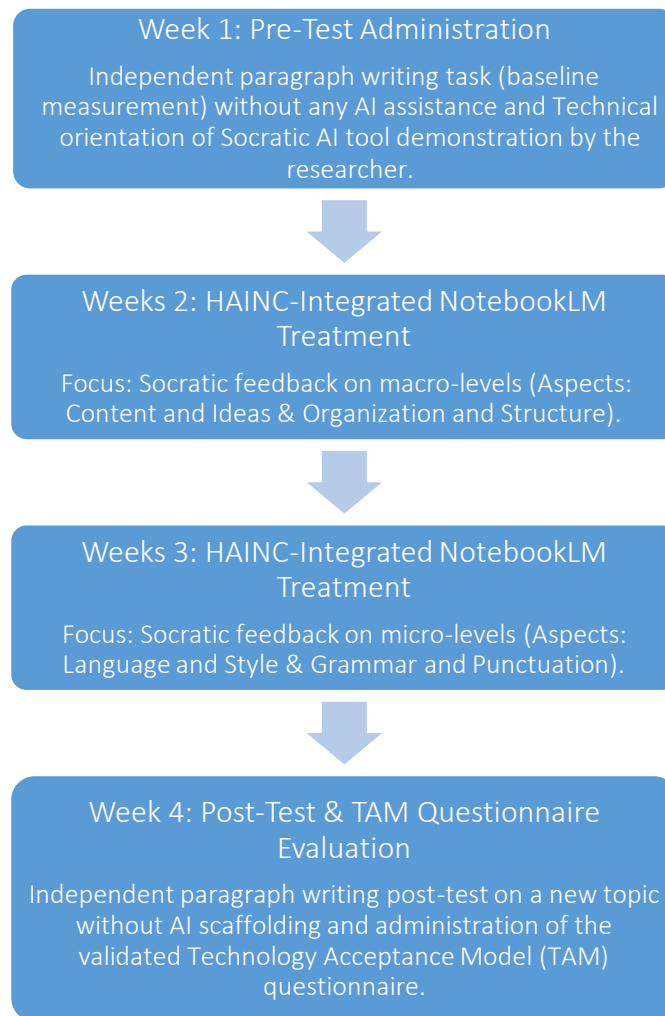


Figure 1. Procedural Flowchart of the Experimental Research Process

3. Result

3.1 The Improvement of Students' Writing Performance

To examine the impact of HAINC-Integrated NotebookLM AI feedback on students' writing performance, a paired sample t-test was conducted by comparing students' pre-test and post-test scores. The writing assessment focused on four components: content and ideas, organization and structure, language and style,

and grammar and punctuation. Each component was individually graded using a standardized 1–5 scale based on the descriptive rubric.

Table 1. Updated Descriptive Statistics of Students' Total Writing Scores

Test	Mean	Standard Deviation (SD)	Variance	Mean Gain Score	Normalized Gain (N-gain)	N
Pre-test	8.17	2.06	4.24	1.26	0.33	23
Post-test	9.43	2.46	6.07		(Moderate)	23

Table 1 shows that the mean score increased from 8.17 in the pre-test to 9.43 in the post-test after the implementation of HAINC-Integrated NotebookLM AI feedback. The standard deviation also shifted from 2.06 to 2.46, reflecting a slightly wider distribution of scores as individual students advanced at varying paces during the intervention. To determine the precise magnitude of this improvement, the Normalized Gain (N-gain) score was calculated based on the maximum threshold of the current score baseline.

The resulting N-gain value of 0.33 falls within the "Moderate" enhancement threshold, indicating a mathematically sound upward progression in students' writing capacities over the four weeks.

To determine whether the improvement was statistically significant, a paired sample t-test was conducted. The result of the analysis is presented in Table 2.

Table 2. Paired Sample t-Test Result

Statistical Analysis	Value
t-value	-3.244
Degree of Freedom (df)	22
Sig. (2-tailed)	0.003
Effect Size (Cohen's d)	0.68 (Medium-to-Large)

The result revealed that the significance value was lower than 0.05 ($0.003 < 0.05$). Therefore, the null hypothesis was rejected, indicating that there was a statistically significant difference between students' pre-test and post-test scores. In other words, the implementation of HAINC-Integrated NotebookLM AI feedback significantly improved students' writing performance.

Furthermore, to quantify the practical strength of this intervention beyond mere statistical significance, an Effect Size analysis was computed utilizing Cohen's d. The analysis yielded a Cohen's d value of 0.68, which denotes a medium-to-large practical effect size. This validates that the observed pedagogical gains were substantial and highly meaningful in a real classroom environment.

To provide a more granular breakdown of these statistical advancements, the overall scores were disaggregated into the four primary writing sub-components. This specific component-wise comparison is delineated in Table 3:

Table 3. Granular Mean Comparisons Across Specific Writing Components (Scale1-5)

Writing Component (Max Score = 5.00)	Pre-Test Mean	Post-Test Mean	Mean Jump	Significance (p-value)
Content and Ideas	2.17	2.52	+0.35	0.012
Organization and Structure	2.43	2.78	+0.35	0.002
Language and Style	2.13	2.26	+0.13	0.041
Grammar and Punctuation	1.43	1.87	+0.44	0.024

As illustrated in Table 3, the most pronounced absolute improvement occurred within the "Grammar and Punctuation" domain (+0.44, p=0.024), followed closely by both "Organization and Structure" (+0.35, p=0.002) and "Content and Ideas" (+0.35, p=0.012). This statistical trend suggests that the Socratic interaction mode of NotebookLM was highly potent in guiding structural arrangement and mechanical accuracy.

Conversely, components like "Language and Style" showed a statistically significant yet more conservative increment (+0.13, p=0.041), verifying that the treatment prioritized intellectual cognitive restructuring and structural synthesis over superficial stylistic changes.

To visually conceptualize the structural leap across these distinct domains on the rubrics' baseline, Figure 2 highlights the comparative milestones between the pre-test and post-test data:

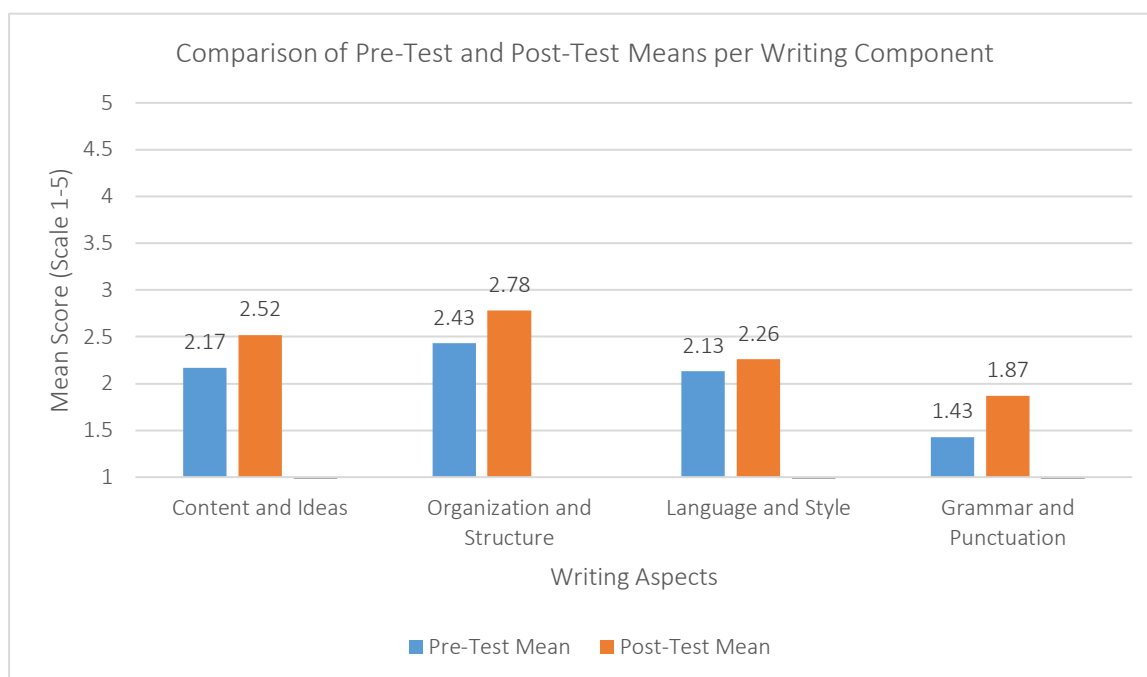


Figure 2. Disaggregated Bar Chart of Pre-Test and Post-Test Component Means

To illuminate the qualitative transformations underlying these empirical measurements, the following text excerpts demonstrate a representative student's iterative negotiation journey utilizing NotebookLM under the HAINC protocol:

Student Excerpt 1: Socratic Dialogue and Student's Text Negotiation Phase (Focus: Content and Ideas)

Initial Pre-Test Draft: "First, they often have problems with grammar and structure. They don't know how to use tenses in sentences."

Human-AI Socratic Interaction Loop (Negotiation Phase):

Student: I want to focus on Content and Ideas first.

NotebookLM Feedback: Great choice! Focusing specifically on Content and Ideas is a fantastic way to make your essay strong and meaningful. I really like how clearly you state the problems students face right from the beginning of your draft.

Let's look closely at **the beginning of your second paragraph, specifically this part: "First, they often have problems with grammar and structure. They don't know how to use tenses in sentences."**

This is a very good start! However, according to Hyland's criteria, strong content needs good "depth of ideas" so the reader can fully understand the problem. Right now, your

idea is clear, but we can make it much deeper by adding a little more detail.

[YOUR TASK]

To give your idea more depth, what specific example could you add here to explain why tenses are confusing, or which specific tenses are usually hard for students?

(Note: Once you think of an answer, please update your draft with your new sentences and provide the revision in the source panel so we can review it together!)

Student: Because we need to use various verbs to indicate specific times, English tenses can be confusing for students, especially because they are very different from Bahasa Indonesia. For example, students may write "I have finished my homework yesterday" instead of "I finished my homework yesterday." This can make students confused about which tense they should use.

NotebookLM Feedback: Wow, that is a brilliant addition! By adding the explanation about the differences with Bahasa Indonesia and providing a specific example of tense misuse, you have significantly improved the **Content and Ideas** of your essay based on Hyland's criteria. This revision gives much better **depth of ideas** to your writing. The positive impact is huge because it helps the reader easily visualize and understand exactly why English tenses are a big challenge for students!

Please make sure to update your draft in the source panel with your new sentences first so I can see your revised text as our proof!

Final Revised Draft: "First, they often have problems with grammar and structure. They don't know how to use tenses in sentences. Because we need to use various verbs to indicate specific times, English tenses can be confusing for students, especially because they are very different from Bahasa Indonesia. For example, students may write "I have finished my homework yesterday" instead of "I finished my homework yesterday." This can make students confused about which tense they should use."

Student Excerpt 2: Adaptive Scaffolding via the Strict Three-Try Rule Protocol (Focus: Grammar and Punctuation)

Initial Pre-Test Draft: "...their writing become simple..."

NotebookLM Feedback (Initial Query / Literal Question): Alright, we can definitely switch our focus to Grammar and Punctuation! Focusing on grammar is a great way to ensure the mechanical accuracy of your writing. I really appreciate your enthusiasm and proactive attitude in wanting to polish different parts of your essay.

However, before we jump into this new topic, please remember to update your draft first and provide the proof of your previous revisions in the source panel. This helps us keep track of your awesome progress!

Once you have updated your text, **let's look at the middle part of your second paragraph, specifically at this phrase: "...their writing become simple..."**.

[YOUR TASK] In English grammar, when you have a singular subject like "writing," what do you need to do to the verb "become" so that they agree with each other? Let me know your thoughts.

Student (Attempt 1 - Tense Misconception): I think we should change it to the past tense? Because they already wrote it. So it should be 'their writing became simple'?

NotebookLM Feedback (Second Cue / Progressive Hinting): That is very critical thinking to consider the timeframe (tense)! However, looking at your paragraph as a whole, you are explaining a general fact that happens in the present, so we should maintain the Simple Present Tense.

In the Simple Present Tense, if the subject is a singular noun (like 'writing' or 'it'), there is a specific letter we usually add to the end of the base verb. Think back, what letter should be added to the word 'become'?

Student (Attempt 2 - Incorrect Guess): Oh, present tense. Do we add '-ing'? Like 'their writing is becoming simple'?

NotebookLM Feedback (Third Cue / Progressive Hinting): Great effort! Adding '-ing' is used for the Continuous Tense (something happening right now). However, because we are using the foundational rules of the Simple Present Tense to state a fact, we do not use '-ing'.

Here is a hint: This rule is exactly the same as when you use the subjects 'He' or 'She' (for example: He plays, She reads). Based on that pattern, what should you do to the verb 'become'?

Student (Attempt 3 - Stalled / Triggering the Rule): I'm still confused. I don't know what to do with the word 'become'. I'm stuck, can you give me a more specific hint?" (Note: The student fails to deduce the correct form after three attempts and explicitly deploys a linguistic threshold phrase signaling cognitive saturation).

NotebookLM Feedback (Execution of the Three-Try Rule / The Pivot Rule): Of course, don't worry, let's look at the rule together! In English, the Subject-Verb

Agreement rule for the Simple Present Tense states that if the subject is singular (such as the word 'writing'), the verb that follows must take an '-s' or '-es' suffix.

Therefore, the verb 'become' must be modified by adding an '-s' at the end. Now, please apply this rule to your draft and rewrite the complete phrase so that it becomes grammatically accurate!

Final Revised Draft: "...their writing becomes simple..."

The qualitative data from both excerpts clearly shows how the HAINC protocol helps students improve their writing through several practical mechanisms. First, the system guides students using progressive prompting rather than simply rewriting the text for them. As seen in Excerpt 1, the AI asks a guiding question to make the student think, which encourages the student to actively expand their ideas by comparing English and Bahasa Indonesia, thereby adding much better depth to their content. Second, the interaction successfully enforces the Three-Try Rule to manage learner progression.

Excerpt 2 demonstrates this rule in action; when the student makes incorrect guesses in their first two attempts, the AI holds back the direct answer and only provides smaller, progressive hints. The AI only explains the underlying Subject-Verb Agreement rule after the student gets completely stuck on the third try, ensuring that the student must still type the final correction manually into their draft.

Furthermore, this interactive loop facilitates an active negotiation of meaning between the human learner and the technology. The students are not passive users who blindly accept automatic proofreading overhauls; instead, they undergo a real discussion or negotiation phase where they must share their own thoughts, try to answer the AI's hints, and test their grammar knowledge before making the final edit. Finally, this method plays a vital role in keeping students motivated during the learning process. The AI consistently uses positive reinforcement and encouraging validation at the beginning of its responses, which prevents students from feeling frustrated when they make mistakes and keeps them engaged until the writing task is fully completed.

The majority of students obtained higher scores in the post-test compared to the pre-test. Several students showed notable improvement, particularly in content organization and grammar accuracy during the revision process. This finding suggests that the Socratic interaction and progressive prompting strategies integrated into the HAINC protocol effectively supported students in revising and developing their writing.

3.2 Students' Technology Acceptance toward HAINC-Integrated NotebookLM AI Feedback

To evaluate the psychological integration and acceptance of NotebookLM as a Socratic writing tool, a 12-item Technology Acceptance Model (TAM) questionnaire was administered across four dimensions: Perceived Usefulness (PU), Perceived Ease of Use (PEOU), Attitude Toward Using AI (ATU), and Behavioral Intention to Use (BI). The diagnostic instruments were statistically validated before interpretation, demonstrating strong internal psychometric properties ($r_{\text{count}} > r_{\text{tabel}}$ of 0.413 for all items, and a Cronbach's alpha coefficient of 0.836, indicating high reliability).

Table 4. Cumulative Distribution and Descriptive Mean of TAM Metrics

TAM Aspect	Items	Cumulative Score	Dimension Mean (Max = 5.00)	Descriptive Level
Perceived Usefulness (PU)	Q1,Q2,Q3	273	3.96	High
Perceived Ease of Use (PEOU)	Q4,Q5,Q6	242	3.50	High
Attitude Towards Using AI (ATU)	Q7,Q8,Q9	263	3.82	High
Behavioral Intention to Use (BI)	Q10	87	3.78	High
Overall Acceptance	TAM Total Items	865	3.77	High Acceptance

As illustrated in Table 4, the cumulative sum of all questionnaire metrics reached 865, yielding an overall baseline mean of 3.76 out of 5.00. This empirical outcome indicates a highly positive disposition toward the integration of AI scaffolding in writing courses. Perceived Usefulness (PU) generated the highest operational mean (3.96), which explicitly clarifies that students considered the structural hints and progressive prompting loop highly effective in boosting their actual writing output. Concurrently, Perceived Ease of Use (PEOU) registered a more conservative mean of 3.51.

This slight divergence is highly consistent with the pedagogical nature of the HAINC protocol; because the AI was strictly restricted from offering immediate answers or direct textual corrections, students encountered an intentional cognitive struggle ("intellectual heavy lifting") which marginally complexified their user-friendliness metrics, yet fundamentally fortified their actual learning gains.

Before analyzing the responses, the questionnaire items were tested for validity and reliability. The validity test showed that all questionnaire items were valid because each correlation value exceeded the r-table value of 0.413. Meanwhile, the reliability test indicated a Cronbach's Alpha coefficient of 0.836, showing that the questionnaire was highly reliable.

Table 5. Questionnaire Validity and Reliability Result

Measurement	Statistical Indicator	Metric Value	Empirical Interpretation
Construct Validity	Pearson Correlation (r-count vs r-table)	r-count > 0.413	All 10 Items are Statistically Valid
Internal Consistency	Cronbach's Alpha	0.836	Highly Reliable Instrument (>0.70)

4. Discussion

4.1 Critical Interpretation of Socratic AI Feedback in Writing

The empirical findings clearly demonstrate that the integration of the HAINC protocol with NotebookLM significantly elevates students' writing performance, particularly in organizing ideas and mastering grammatical accuracy. Unlike conventional pedagogical setups where instructional feedback is predominantly unidirectional, the collaborative framework tested in this study triggered a dynamic cognitive engagement. The notable transformation from fragmented sentences during the pre-test to cohesive, structurally sound arguments stems directly from the Socratic interaction mechanism.

Instead of offering immediate corrections, the AI functioned as a pedagogical scaffold that intentionally delayed gratification, forcing students to navigate their own linguistic gaps. By withholding explicit answers, the progressive prompting cycle forced students to engage in deep metalinguistic reflection based on the core analytical domains (Hyland, 2019). When students were prompted to locate errors independently, they underwent a process of cognitive restructuring, mapping their implicit linguistic knowledge onto explicit corrective actions. As emphasized by Chen (2025), AI feedback becomes profoundly effective when students are actively involved in evaluating and negotiating AI-generated suggestions rather than passively consuming automated choices.

This cognitive leap is heavily anchored in Vygotsky's Zone of Proximal Development (ZPD), where the AI serves as an adaptive scaffolding agent to foster autonomous learning (WHW Hamedi et al., 2025). The "Three-Try Rule" inherent in the operational manual operationalized this scaffolding by systematically managing the students' cognitive load. By addressing only one major structural or

grammatical anomaly at a time, the system prevented the cognitive paralysis often associated with exhaustive, unstructured feedback. This structured approach directly supports Cognitive Load Theory (Ahmed & Uddin, 2025), which posits that AI-assisted scaffolding reduces an L2 learner's cognitive burden during complex tasks by breaking down the revision process into manageable, micro-level stages.

Consequently, students did not merely receive passive corrections; they actively negotiated meaning with the digital tutor using robust interactional strategies (WEN Qiufang, 2024). This negotiated framework aligns closely with the core principles of Self-Regulated Learning (SRL), wherein students are systematically prompted to monitor, evaluate, and adjust their writing behavior. By positioning learners as active decision-makers rather than passive recipients of automatic proofreading (Al-Kadi, 2025), the HAINC protocol effectively balances technological assistance with human autonomy (Soori et al., 2025).

Therefore, the statistical improvements recorded in the results are not arbitrary artifacts of technology usage, but are reflections of deep-seated learner autonomy and heightened metalinguistic awareness cultivated through deliberate, repetitive problem-solving cycles.

4.2 Comparative Analysis: NotebookLM vs. Grammarly and ChatGPT

To appreciate the pedagogical uniqueness of the current intervention, a comparative analysis must be drawn between the source-based Socratic feedback of NotebookLM and alternative AI architectures, such as automated proofreading tools (e.g., Grammarly) and direct generation engines (e.g., ChatGPT). Automated writing evaluation (AWE) tools like Grammarly operate primarily on a surface-level error correction paradigm (Zhai, 2026).

While highly efficient for mechanical rectifications, such tools foster a passive editing experience where students blindly accept automated stylistic overhauls without internalizing the underlying grammatical principles required in academic writing (Hyland, 2019). Conversely, large language models like standard ChatGPT often exhibit an over-permissive output generation tendency. When prompted with a flawed text, ChatGPT frequently rewrites the entire paragraph instantly, stripping the student of the opportunity to undergo the essential intellectual struggle required for deep writing development.

The empirical literature confirms that while generic AI-generated feedback positively affects EFL writing quality by providing immediate support (Alghannam, 2025) and broadly enhances writing achievement across diverse contexts (Fleckenstein et al., 2023), These systems frequently default to automated proofreading. NotebookLM, within the specified HAINC parameters, breaks away from these conventional models through its retrieval-augmented generation (RAG) technology and strict interaction constraints. It operates not as an editor, but as a dialogic partner.

While Grammarly bypasses the student's cognitive processing and ChatGPT replaces it entirely, NotebookLM enforces intense intellectual heavy lifting. It isolates structural discrepancies and cross-references them directly against pre-loaded criteria (Hyland, 2019), compelling the user to articulate and execute the necessary revisions manually within their dual worksheets. This source-grounded approach expands the current boundary of computer-assisted language learning (CALL) by transforming AI from a text-generating crutch into a robust Socratic learning partner in the Indonesian EFL context.

4.3 Mitigation of Inherent AI Limitations and Ethical Challenges

Despite the positive empirical output, large language models possess inherent technical vulnerabilities, notably data hallucinations and text biases, which pose significant risks in educational contexts. Hallucination—the phenomenon where an AI generates statistically plausible but factually or grammatically incorrect data—can severely mislead novice L2 learners who lack the linguistic maturity to verify AI assertions. In this study, this limitation was systematically mitigated through the deployment of RAG-grounded constraints. By strictly anchoring NotebookLM's analytical parameters to the uploaded 'Tutor Writing Manual' adapted from Chen (2025), the probability of erratic or ungrounded feedback was successfully constrained. The RAG framework forced the AI to generate queries derived strictly from authorized instructional principles, ensuring a high degree of psychometric reliability.

Furthermore, the integration of the HAINC protocol actively resolved the overarching ethical dilemma of academic overreliance. Critics of generative AI in education frequently argue that automated feedback systems encourage academic dishonesty and diminish critical thinking (Garcia et al., 2025). However, the HAINC protocol reverses this dynamic by transforming the user experience from consumption to critical negotiation. Because the master prompt structurally forbids the system from executing direct rewrites, students cannot use the AI to bypass the writing process.

The student must manually edit their drafts based on the conceptual hints provided. This confirms that AI technologies can promote highly meaningful formative feedback only when learners critically interact with the system rather than relying solely on automated generation (Burner et al., 2025). Therefore, this framework serves as an ethical blueprint for AI integration, proving that when guided by a disciplined pedagogical protocol, technology can be weaponized to combat overreliance rather than induce it.

4.4 Pedagogical Implications, Facilitator Shifts, and Curriculum Sustainability

The successful execution of this model carries extensive pedagogical implications for modern EFL education, signaling a structural shift in classroom dynamics. The traditional role of the writing lecturer as the absolute, solitary evaluator is no longer sustainable or efficient in high-enrollment contexts. Under this new paradigm, the instructor's role shifts from a mechanical proofreader to an instructional designer and prompt engineer. In the collaborative classroom, the human educator is liberated from basic mechanical corrections and can instead focus on higher-order rhetorical counseling and managing affective factors, such as writing anxiety and motivation.

To ensure the future institutional viability and sustainability of this approach in higher education, institutional curriculums must undergo proactive updates. Rather than banning or uncritically adopting AI tools, universities should integrate structured AI-literacy modules directly into the syllabus of academic writing courses, drawing on modern acceptance frameworks (Davis & Granić, 2024). The high technology acceptance observed across all TAM dimensions in this study underscores that students perceive localized AI feedback as highly practical, user-friendly, and simple to adopt (Khojasteh et al., 2025; Rozaimie & Mumin, 2025)

Contextually, this positive trend deeply correlates with behavioral traits among students at IAIN Parepare. As documented by Noviasmy et al. (2023) Students at this institution exhibit a strong attachment to mobile digital devices. This baseline digital behavior is further explained by Usman et al. (2023), who stated that the pervasive integration of smartphones into students' daily routines, paired with high intrinsic motivation, greatly facilitates the acceptance of digital learning tools that deliver immediate validation.

In high-enrollment Indonesian higher education setups where students frequently struggle to receive immediate, individualized feedback from lecturers due to time constraints, establishing permanent RAG environments pre-loaded with university-approved rubrics becomes a pedagogical necessity rather than a technological luxury. By institutionalizing protocols like HAINC, higher education institutions can transition from manual, time-consuming evaluation methods toward a sustainable, technology-enhanced pedagogical ecosystem.

4.5 Limitations of the Study

While this study provides valuable empirical insights into AI-scaffolded writing instruction, several inherent limitations must be recognized. First, the research design utilized a one-group pre-test post-test framework without a traditional control group. Consequently, while the statistical advancements are highly significant, the absence of a control cohort limits the ability to isolate the effects of NotebookLM completely from other external variables, such as natural

maturation or concurrent linguistic exposure.

Second, the sample size was restricted to 23 EFL students within a single institutional context at IAIN Parepare, which restrains the broad generalization of the findings to larger, more diverse student bodies.

Third, the intervention period was constrained to a strict four-week cycle, measuring immediate post-treatment performance. This timeframe leaves the long-term retention of these independent writing enhancements unexamined, necessitating future longitudinal tracking. Finally, the reliance on self-reported psychological metrics via the TAM questionnaire introduces potential social desirability bias, where students might overstate their positive attitudes toward the technology.

Future research endeavors should employ randomized controlled trials (RCTs), incorporate delayed post-tests to track long-term retention, and integrate multi-institutional sample selections to expand the empirical validity of Socratic AI feedback in language learning ecosystems.

5. Conclusion

This study investigated the impact of HAINC-Integrated NotebookLM AI feedback on EFL students' writing performance at IAIN Parepare. The findings demonstrated that the implementation of NotebookLM as a Socratic AI tutor significantly improved students' writing achievement, particularly in content development, organization, language use, and grammatical accuracy. The statistical analysis revealed a significant difference between the pre-test and post-test scores, indicating that AI-assisted feedback integrated with the HAINC protocol effectively supported students during the writing revision process. Through progressive prompting and the "Three-Try Rule," students were encouraged to engage more actively, critically, and autonomously in evaluating and revising their writing.

Furthermore, the findings of the Technology Acceptance Model (TAM) questionnaire indicated that students showed a high level of acceptance toward the implementation of NotebookLM AI feedback. Students perceived the system as useful, accessible, and beneficial in facilitating writing activities and improving learning efficiency. The positive responses toward perceived usefulness, perceived ease of use, attitude toward using AI, and behavioral intention suggest that AI-assisted feedback systems have strong potential to support EFL writing instruction in higher education contexts. The integration of Socratic AI interaction also contributed to students' confidence and reflective learning during the revision process.

Despite these positive findings, this study was limited by the small sample size and the use of a pre-experimental design without a control group. Therefore, the results cannot be generalized broadly across different educational contexts. Future research is recommended to employ experimental or quasi-experimental designs

involving larger and more diverse participants to obtain more comprehensive findings. In addition, future studies may further explore the implementation of HAINC-integrated AI feedback in other language skills, including speaking, reading, and academic research writing, as well as investigate the long-term impact of AI-assisted Socratic learning on students' critical thinking and learner autonomy.

6. Acknowledgement

The author would like to express sincere gratitude to Institut Agama Islam Negeri (IAIN) Parepare, especially the English Education Department of the Tarbiyah Faculty, for providing academic support and research facilities throughout this study. The author also extends appreciation to all students who participated in this research for their valuable contribution and active participation during the implementation of the study. Special thanks are addressed to the supervisors and examiners for their constructive feedback, insightful suggestions, and continuous guidance in improving the quality of this research article.

7. References

- Ahmed, M. R., & Uddin, M. M. (2025). Multimedia Integration to Enhance Interactivity: A Cognitive Load Theoretical Approach in American Poetry Instruction. *International Journal of Humanities Education*, 23(1), 125–148. <https://doi.org/10.18848/2327-0063/CGP/v23i01/125-148>
- Al-Kadi, A. (2025). Fostering a 'Human AI' Approach for Evaluating Students' Writing in English. *Studies in Linguistics, Culture and FLT*, 13(1), 140–159. <https://doi.org/10.46687/VFBZ9792>
- Alghannam, M. S. M. (2025). Artificial Intelligence as a Provider of Feedback on EFL Student Compositions. *World Journal of English Language*, 15(2), 161–173. <https://doi.org/10.5430/wjel.v15n2p161>
- Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: a quantitative study. *Frontiers in Education*, 10. <https://doi.org/10.3389/feduc.2025.1614673>
- Bahari, A. (2026). Evaluating the Effectiveness of AI-Driven Approaches on EFL Learners' Expository Writing Skills. *TESOL Journal*, 17(1). <https://doi.org/10.1002/tesj.70099>
- Burner, T., Lindvig, Y., & Wærness, J. I. (2025). "We Should Not Be Like a Dinosaur" — Using AI Technologies to Provide Formative Feedback to Students. *Education Sciences*, 15(1). <https://doi.org/10.3390/educsci15010058>
- Chen, Q. (2025). *Students' Perceptions of AI-Powered Feedback in English Writing: Benefits and Challenges in Higher Education*. 00(June), 1–11. <https://doi.org/10.47852/bonviewIJCE52025580>

- Davis, F. D., & Granić, A. (2024). Evolution of TAM. In *The technology acceptance model: 30 years of TAM* (pp. 19–57). Springer.
- Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: a multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence*, 6. <https://doi.org/10.3389/frai.2023.1162454>
- Garcia, M. B., Rosak-szyrocka, J., & Ramazan, Y. (2025). Rethinking Educational Assessment in the Age of Generative AI: Actionable Strategies to Mitigate Academic Dishonesty. *Pitfalls of AI Integration in Education: Skill Obsolescence, Misuse, and Bias*. [https://doi.org/https://doi.org/10.4018/979-8-3373-0122-8.ch001](https://doi.org/10.4018/979-8-3373-0122-8.ch001)
- Hyland, K. (2019). *Second language writing*. Cambridge university press.
- J Eppendi, A. A. (2025). *Fostering Learner Autonomy in EFL Classroom: The Impact of New Learning Guidelines and Relational Pedagogy*. 215–236.
- Johnson, R. B., & Christensen, L. B. (2024). *Handbook*. Sage publications.
- Khojasteh, L., Karimian, Z., Nasiri, E., Kafipour, R., & Farahmandi, A. Y. (2025). Artificial intelligence and academic writing questionnaire (AI-AWQ): development and validation among medical students' experiences using exploratory factor analysis. *BMC Medical Education*, 25(1). <https://doi.org/10.1186/s12909-025-08288-z>
- Noviasmy, Y., Dalle, A., & Hasanah, N. (2023). Applying Quizizz Application as an Assessment Tool for EFL Students. *Inspiring: English Education Journal*, 6(1), 12–22.
- Pertiwi, P. C., Segoh, D., & Rohmadhani, A. (2025). Artificial Intelligence in Academic Environments : Reducing or Increasing Foreign Language Anxiety ? *Inspiring: English Education Journal*, 8(1), 100–131. <https://doi.org/DOI:10.35905/inspiring.v8i1.13009>
- Rozaimie, F., & Mumin, M. A. (2025). ChatGPT Integration in Writing Development: Student Experiences and Perspectives. *World Journal of English Language*, 15(8), 74–83. <https://doi.org/10.5430/wjel.v15n8p74>
- Ruby, M., Chen, T., Zhou, T., & Hsieh, G. (2026). PaperTok: Exploring the Use of Generative AI for Creating Short-form Videos for Research Communication. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain* (Vol. 1, Issue 1). Association for Computing Machinery. <https://doi.org/10.1145/3772318.3790553>
- Santos, É. M. B., Hirayama, D., & Costa, T. B. (2025). *Artificial Intelligence in Engineering Learning: How Do Engineering Students Use It?* 163–170. <https://doi.org/10.5281/zenodo.15850339>
- Soori, A., Khojasteh, L., & Javed, F. (2025). Comparing Teacher E-Feedback, AI

- Feedback, and Hybrid Feedback in Enhancing EFL Writing Skills. *Technology in Language Teaching and Learning*, 7(3).
<https://doi.org/10.29140/tltl.v7n3.102626>
- Tufino, E. (2025). NotebookLM as a Socratic physics tutor: Design and preliminary observations of a RAG-based tool. *ArXiv Preprint ArXiv:2504.09720*.
- Usman, U., Fakhruddin, Z., Hardiyanti, H., Adam, Z., Kadaruddin, K., & Rahmani, B. (2023). Developing EFL Teachers' Competence in Designing Learning Materials through Electronic English Book Design Training. *Jurnal Inovasi Pembelajaran*, 2(9), 140–153.
<https://doi.org/https://doi.org/10.22219/jinop.v9i2.25287>
- WEN Qiufang, L. M. (2024). Human-AI interactive negotiation competence: ChatGPT and foreign language education. *Foreign Language Teaching and Research*, 56(2), 286–296.
- WHW Hamedi, FDA Ali, W. A. (2025). AI as A Digital Scaffold : an Integrative Review of Vygotsky's Zone of Proximal Development in Modern Education. *International Journal of Modern Education (IJMOE)*, 7(26), 579–589.
<https://doi.org/10.35631/IJMOE.726038>
- Zhai, Y. (2026). *Automated written corrective feedback : fostering writing accuracy , self-regulation , and motivational dynamics in EFL learners*.
- Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers and Education*, 143.
<https://doi.org/10.1016/j.compedu.2019.103668>