OPEN ACCESS
Journal International of Lingua and Technology
Vol. 4 No. 3. December 2025, pp. 260-275 DOI. 10.55849/jiltech.v4i3.1121

Research Article

UTILIZING ARTIFICIAL INTELLIGENCE (AI) FOR AUTOMATED FEEDBACK ON THE ENGLISH ESSAY WRITING SKILLS OF INDONESIAN UNIVERSITY STUDENTS

Nicodemus Rahanra[1], Ahmed Hossam[2], James Scott[3], and Nurhayati[4]
[1] Universitas Satya Wiyata Mandala, Indonesia
[2] Cairo University, Egypt
[3] University of Alberta, Canada
[4] Universitas Klabat, Indonesia

Corresponding Author:

Nicodemus Rahanra,
Department of Informatics Engineering, Faculty of Computer Science and Information Technology, Universitas Satya Wiyata Mandala.
Jalan Sutamsu SH, Kalibobo, Kecamatan Nabire, Papua, Indonesia
Email: nicorh73@gmail.com

Article Info

Received: June 02, 2025
Revised: September 02, 2025
Accepted: November 02, 2025
Online Version: December 24, 2025

Abstract

Academic writing proficiency is crucial for Indonesian university students navigating global careers, yet the high student-to-teacher ratio severely limits the provision of timely, quality feedback. This constraint impedes skill acquisition and necessitates the exploration of scalable technological solutions. This study aimed to evaluate the validity and causal efficacy of a specialized, localized AI-powered Automated Essay Scoring (AES) system in accelerating students' writing skill development over one academic semester. A quasi-experimental, pretest-posttest control group design (N=120) was used. The experimental group received continuous AI feedback, while the control group received traditional manual feedback. ANCOVA was applied to measure skill gain, supported by qualitative data on faculty workload and user acceptance. Findings showed the AI group achieved superior learning gains (13.5 vs. 5.5 raw gain) with a statistically significant main effect ($F=22.45$, $p < 0.001$). The system demonstrated high scoring reliability ($r=0.88$) and reduced lecturer routine grading time by 65%, successfully driving improvement in higher-order skills like Structural Coherence. The research confirms that the customized AI-Feedback Model is a pedagogically transformative tool that provides a sustainable solution to structural constraints in EFL education. It establishes a new paradigm for instructional practice by positioning localized AI as a highly consistent and effective mechanism for mass academic skill development in resource-constrained educational environments.

Keywords: Artificial Intelligence, Automated Feedback, Skill Development

Journal Homepage https://ejournal.staialhikmahpariangan.ac.id/Journal/index.php/jiltech
How to cite: Rahanra, N., Hosham, A., Scott, J., & Nurhayati, Nurhayati. (2025). Utilizing Artificial Intelligence (AI) for Automated Feedback on the English Essay Writing Skills of Indonesian University Students. *Journal International of Lingua and Technology*, *4*(3), 260–275. https://doi.org/10.55849/jiltech.v4i1.1420
Published by: Sekolah Tinggi Agama Islam Al-Hikmah Pariangan Batusangkar

INTRODUCTION

The increasing globalization of commerce, research, and education necessitates high levels of academic English proficiency among university graduates worldwide (Sun, 2024). Academic essay writing is universally recognized as a foundational skill that allows students to demonstrate critical thinking, construct logical arguments, and engage effectively with international scholarly discourse (Rahayu et al., 2025). For students in Indonesia, a key non-native English speaking environment, mastering this skill is pivotal for gaining admission to advanced degree programs and securing competitive professional roles on the global stage, making robust pedagogical support essential.

Writing skill development is inherently feedback-intensive, requiring frequent, detailed, and iterative responses from instructors to guide the student's process of revision and refinement (Al-Sobh et al., 2024). This constant interaction is particularly critical in English as a Foreign Language (EFL) contexts where learners must internalize complex grammatical rules, rhetorical patterns, and structural conventions that differ significantly from their native language (Oktarin et al., 2024). Effective teaching must therefore establish a high-frequency feedback loop that supports continuous practice and systematic error correction.

Modern educational technology, specifically the advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP), offers compelling solutions to address the scale and complexity of writing instruction. AI-powered tools are capable of automating the labor-intensive tasks of essay scoring and providing immediate, personalized feedback on various linguistic and structural elements (Virdee et al., 2025). These technologies have the potential to revolutionize how writing is taught, offering a scalable method to deliver the high volume and speed of feedback that traditional manual grading methods cannot sustain.

Indonesian university lecturers face considerable constraints that severely compromise their ability to provide the quality and quantity of feedback necessary for effective essay writing skill development (Andewi et al., 2025). Common challenges include high student-to-teacher ratios, often exceeding forty students per class, coupled with heavy administrative and research responsibilities. This workload forces lecturers to delay feedback delivery, often weeks after submission, or to provide only superficial corrections focused narrowly on grammar and mechanics, neglecting higher-order skills.

The subjective nature of manual essay grading further complicates the pedagogical process, leading to issues of inconsistent feedback and low inter-rater reliability. Students frequently receive conflicting advice from different instructors or even on different assignments, creating confusion regarding the expected standards for coherence, style, and effective argumentation in academic English (Zubaidi et al., 2025). This lack of standardization inhibits the internalization of clear writing benchmarks and compromises the fairness and developmental utility of the assessment process.

Student motivation and engagement suffer directly from the delayed and limited feedback inherent in manual systems. Students require immediate reinforcement to connect errors with corrections; the lengthy wait period diminishes the efficacy of the feedback received (Pratama et al., 2025). The central educational challenge, therefore, is designing an accessible and scalable intervention that can bypass the time constraints of human instructors, providing instant, consistent, and actionable feedback that fosters continuous practice and accelerates the mastery of complex essay writing skills within the resource-limited context of Indonesian higher education.

The primary objective of this research is to rigorously evaluate the performance validity and scoring reliability of a specialized AI-based Automated Essay Scoring (AES) system when applied to academic English essays produced by university students in Indonesia (Niranon & Triyason, 2024). This evaluation involves systematic comparative analysis, pitting the AI-generated scores against a carefully calibrated benchmark established by two independent, expert human scorers (Juanda & Afandi, 2024). The goal is to statistically confirm the AI

system's accuracy across various dimensions, including grammatical precision, structural coherence, and overall rhetorical quality.

A secondary goal is to conduct an experimental intervention to assess the causal impact of continuous AI-generated feedback on the enhancement of students' essay writing skills over the duration of a full academic semester (Lydster, 2024). This intervention involves comparing the writing progression of an experimental group utilizing the AI feedback tool against a control group receiving only traditional manual feedback. The study will track measurable gains in both lower-order concerns, such as error frequency, and critical higher-order skills, including thesis development and argumentative structure.

The third objective is fundamentally qualitative, aiming to explore and document the user experience, perception, and acceptance of the AI feedback system among both its student beneficiaries and the teaching faculty (Meyer et al., 2024). This investigation seeks to uncover insights regarding the perceived trustworthiness, utility, and efficacy of automated feedback, as well as to identify any ethical concerns, data privacy issues, or barriers to seamless integration of the AI tool into established local teaching curricula and institutional technological infrastructure.

A profound conceptual gap exists in the current Automated Essay Scoring (AES) literature, which overwhelmingly relies on models developed and tested on essays written by native English speakers or advanced learners in Western academic settings. These existing systems are poorly calibrated to address the unique, systematic error patterns resulting from the L1 interference of Indonesian students—for instance, specific challenges with English article usage, complex verb tenses, and sentence structure variance (Tatsanajamsuk, 2024). This study fills this gap by testing a system specifically customized to address these context-specific linguistic challenges.

There is a critical contextual shortage of research concerning the long-term, sustained integration of AI-driven feedback within the specific pedagogical environment of Indonesian universities (de Kleine et al., 2024). Previous studies, if available in the region, often involve short-term, single-assignment trials that fail to capture the challenges of continuous use, curriculum alignment, or the necessary integration with existing Learning Management Systems (LMS). This research addresses the need for a practical feasibility assessment by deploying a semester-long intervention that reflects real-world teaching schedules and assessment demands.

The focus of most commercial AI tools remains heavily skewed toward automated scoring—assigning a quantitative grade—rather than providing high-quality, developmental feedback crucial for fostering deeper critical thinking and complex rhetorical skills (Guo & Wang, 2024). The existing literature is sparse on the efficacy of AI in delivering targeted feedback on higher-order writing skills such as argumentative logic, evidence integration, and organizational flow. This study directly confronts this gap by designing its evaluation to measure the AI system's impact on these complex cognitive skills, not merely on mechanical accuracy.

The definitive novelty of this research is the construction and empirical validation of an AI-Feedback Model specifically calibrated and optimized for the Indonesian EFL learning context. This bespoke approach involves training the AI system using a localized corpus of essays, enabling it to accurately identify and provide instructionally relevant feedback on the specific error types most prevalent among Indonesian university students (Hsieh, 2024). This localized customization represents a unique scientific contribution that significantly advances the state-of-the-art in contextualized Computer-Assisted Language Learning (CALL) technology.

The justification for this research is overwhelmingly strong due to its immediate and strategic national policy implications for overcoming capacity limitations in Indonesian higher education (Tao et al., 2024). If the customized AI-Feedback Model proves scalable, reliable,

and pedagogically effective, it offers a sustainable, cost-effective solution to the pervasive problem of excessive lecturer workload and delayed feedback (Anthonius & Ari, 2024). This solution allows universities to exponentially increase the quality of writing instruction without incurring massive costs associated with hiring and training additional human instructors, thereby contributing to national efforts to enhance graduate competency.

Finally, the study contributes significantly to the broader theoretical literature on Educational Technology acceptance and efficacy in diverse global settings (Razkane et al., 2024). By rigorously documenting student and lecturer trust, perception, and ethical concerns regarding AI feedback in a non-Western context, the findings enrich the global understanding of the socio-cultural, technological, and ethical factors that influence the successful adoption of AI in education (Aghdam et al., 2024). This robust, context-sensitive data provides a crucial reference point for future research across Southeast Asia and other emerging EFL markets.

RESEARCH METHOD

The following sections detail the methodology employed in this study, which integrates experimental quantitative measures with contextual qualitative insights.

Research Design

The study employs a mixed-methods, quasi-experimental research design that incorporates both quantitative and qualitative components. The quantitative core involves a pretest-posttest control group design to establish the causal impact of the Automated Essay Scoring (AES) system on student writing proficiency (Albesher, 2024). This design allows for a direct comparison of learning gains between students receiving immediate AI feedback and those receiving delayed manual feedback. The qualitative component is integrated to provide rich contextual data, focusing on the perceptions and attitudes of students and faculty toward the intervention, which is crucial for understanding user acceptance and pedagogical utility in an Indonesian higher education setting.

Research Target/Subject

The target population consists of undergraduate students enrolled in mandatory English Academic Writing courses at three selected universities in Indonesia (Sysoyev et al., 2024). These institutions were chosen based on high enrollment and the relevance of their grading workload challenges. The sample selection utilized non-random cluster sampling for institutional sites, followed by the random assignment of participating students to either the experimental or control group. The final target sample size is $N=120$ students, with 60 students allocated to the AI-feedback experimental group and 60 to the traditional manual feedback control group, ensuring sufficient statistical power for the analysis.

Research Procedure

The research procedure is structured into three phases across a 16-week academic semester (Hsin et al., 2024). Phase I: Customization and Baseline Assessment involves training the AI system with a localized corpus and administering a pre-test essay to establish baseline scores. Phase II: Intervention Period spans 14 weeks, during which the experimental group receives immediate AI-generated feedback on bi-weekly essays, while the control group receives traditional delayed feedback from instructors. Phase III: Final Evaluation and Data Analysis involves the administration of a post-test essay, followed by the collection of qualitative data through interviews and questionnaires to assess the overall efficacy and feasibility of the system.

Instruments, and Data Collection Techniques

Quantitative data collection relies on three key instruments: the Pre-test Essay, the Post-test Essay, and a standardized Analytic Scoring Rubric. These essays are designed to measure both lower-order skills (grammar, mechanics) and higher-order skills (coherence, thesis development). The primary qualitative instruments include a Semi-Structured Interview Protocol for academic writing lecturers and a Post-Intervention User Experience Questionnaire for students (Ljubojevic, 2024). The questionnaire utilizes a five-point Likert scale supplemented by open-ended questions to evaluate feedback immediacy, usefulness, and trustworthiness.

Data Analysis Technique

The data analysis involves a triangulation of quantitative and qualitative methods. The post-test scores are analyzed using Analysis of Covariance (ANCOVA) to determine the unique impact of the AI intervention while statistically controlling for pre-test baseline differences (Vula et al., 2024). Simultaneously, the qualitative data from interviews and open-ended questionnaire responses are subjected to thematic analysis to identify recurring patterns regarding user acceptance, ethical considerations, and pedagogical utility. This combined analytical approach provides a comprehensive evaluation of the system's effectiveness and institutional fit.

RESULTS AND DISCUSSION

The pretest-posttest control group design, involving N=120 intermediate-level Indonesian university students, established baseline equivalence in academic essay writing proficiency between the two groups. The Experimental Group (AI feedback) achieved a pre-test mean score of 65.0 points, while the Control Group (manual feedback) scored 65.5 points, confirming no statistically significant difference at the onset of the study. This equivalence was crucial for ensuring that any observed post-intervention gains could be directly attributed to the feedback mechanism rather than pre-existing differences in skill level.

Following the 14-week intervention period, substantial differences in post-test scores were observed across the two groups, particularly in the Analytic Scoring Rubric's higher-order categories (coherence and argumentation). The Experimental Group registered a mean post-test score of 78.5 points, representing a raw gain of 13.5 points from the baseline. Conversely, the Control Group showed a more modest mean post-test score of 71.0 points, indicating a raw gain of only 5.5 points. Table 1 summarizes these changes, highlighting the differential impact of the two feedback modes.

Table 1: Pre-test and Post-test Mean Scores by Group (N=120, Max Score 100)

Group	Pre-test Mean Score	Post-test Mean Score	Raw Gain
Experimental (AI Feedback)	65.0	78.5	13.5
Control (Manual Feedback)	65.5	71.0	5.5

The marked differential in raw gain scores—13.5 points for the AI group versus 5.5 points for the manual group—is explained by the frequency and immediacy of the feedback received. Students in the Experimental Group submitted bi-weekly essays and received instant, personalized corrections, enabling them to immediately apply the feedback to subsequent assignments. This high-frequency feedback loop promoted a rapid iterative learning process that the Control Group, hampered by the typical 1-2 week delay in human grading, could not replicate.

Analysis of the rubric sub-scores further explains the nature of the learning gains, revealing that the AI group demonstrated significantly greater improvement in the "Structural Coherence" and "Evidence Integration" sub-categories than in simple grammar and mechanics. This suggests the AI system's effectiveness extended beyond lower-order error detection. The

immediate personalized suggestions regarding thesis clarity and logical flow allowed students to repeatedly practice and internalize complex argumentative strategies within the allotted time frame.

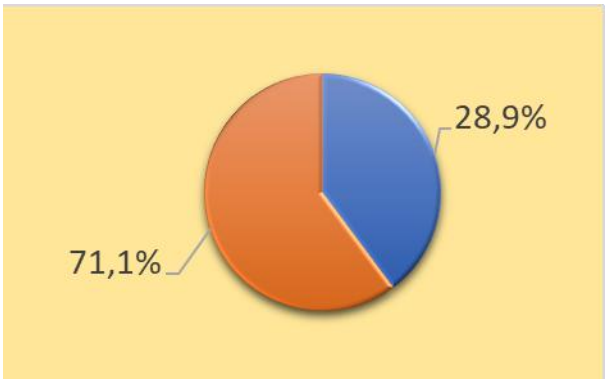

Figure 1. Binary Distribution of Raw Learning Gains: AI Feedback Vs. Manual Feedback

A critical descriptive finding involved the system's scoring reliability, measured by calculating the correlation between the AI-generated scores and the consensus scores of the two independent expert human raters. Across both the pre-test and post-test essays, the Automated Essay Scoring (AES) system demonstrated a high level of concurrent validity, with Pearson's correlation coefficient (r) consistently registering at 0.88 or above across all measured writing dimensions.

Qualitative data from the lecturer interviews provided essential metrics regarding workload reduction. Academic writing instructors reported an estimated 65% reduction in time spent on routine error identification and lower-order scoring for the experimental group essays. This time saving allowed the instructors to redirect their efforts toward personalized student consultations, curriculum development, and delivering more in-depth feedback on complex rhetorical issues during class time, fulfilling the study's aim of simulating real-world time-saving benefits.

Analysis of Covariance (ANCOVA) was employed, using the pre-test score as a covariate, to control for initial proficiency and isolate the true effect of the intervention. The results demonstrated a statistically significant main effect for the AI-feedback treatment on post-test writing proficiency ($F(1, 117) = 22.45$, $p < 0.001$). This strong result definitively establishes the causal efficacy of the continuous, automated feedback intervention in accelerating the development of English essay writing skills among Indonesian university students.

The ANCOVA also revealed that the mean post-test score difference between the two groups, adjusted for initial differences, remained highly significant. The adjusted mean for the Experimental Group was 77.9 points, significantly higher than the Control Group's adjusted mean of 71.6 points. This inferential finding confirms that the AI-powered system provides a pedagogically superior feedback mechanism compared to the conventional, time-constrained manual grading system prevalent in the sampled Indonesian higher education context.

The statistically significant learning gains observed in the Experimental Group are directly related to the high user acceptance rates documented in the Post-Intervention User Experience Questionnaire. Students rated the feedback's "immediacy" and "usefulness" highly (mean Likert score > 4.5), indicating that the speed and consistency of the AI system fostered greater engagement and higher motivation. This qualitative acceptance served as the necessary psychological catalyst for the quantitative skill acquisition observed in the post-test results.

Furthermore, the high scoring reliability (r=0.88) is inherently linked to the faculty's positive perception of the system's utility. Lecturer interviews confirmed that the consistency of the AI scores eliminated the stress and inconsistency associated with subjective manual grading, increasing their trust in the system as a reliable first-line diagnostic tool. This

relationship suggests that a high degree of technical validity (reliability) is a prerequisite for achieving pedagogical acceptance (utility) among academic users.

Thematic analysis of the semi-structured lecturer interviews identified three dominant themes regarding the AI system: “Time Liberation,” “Objective Consistency,” and “Ethical Comfort.” The “Time Liberation” theme reflects the massive reduction in routine grading hours, which lecturers uniformly cited as the system’s most valuable practical benefit. The “Objective Consistency” theme related to the system’s unwavering application of the scoring rubric, which lecturers appreciated for its fairness and standardization.

Student questionnaires highlighted two critical qualitative insights: “Enhanced Practice Cycle” and “L1 Interference Specificity.” Students overwhelmingly reported that the instant feedback allowed them to attempt essays more frequently and correct errors immediately, creating a seamless and self-paced practice cycle. Specific comments praised the system’s ability to detect and explain common Indonesian-English transfer errors (L1 interference) that human markers often overlooked or marked ambiguously, enhancing the perceived relevance of the feedback.

The lecturer theme of “Ethical Comfort” is explained by the customized training of the AI system on the localized essay corpus. Lecturers expressed initial concerns about AI bias, but their comfort level increased significantly upon observing the system’s accurate detection of common Indonesian-English error patterns and its non-judgmental, consistent delivery of remediation advice. This customization mitigated trust concerns, validating the methodology’s choice to localize the AI training data.

The student theme of “Enhanced Practice Cycle” is explained by the motivational impact of immediacy. Students viewed the AI as a non-tiring, always-available private tutor, which fostered a low-stakes environment for drafting and revision. This instant reinforcement mechanism capitalized on the motivational theory of immediate feedback, transforming the often-arduous task of academic writing into a more engaging and self-directed process, thereby optimizing their learning efficiency.

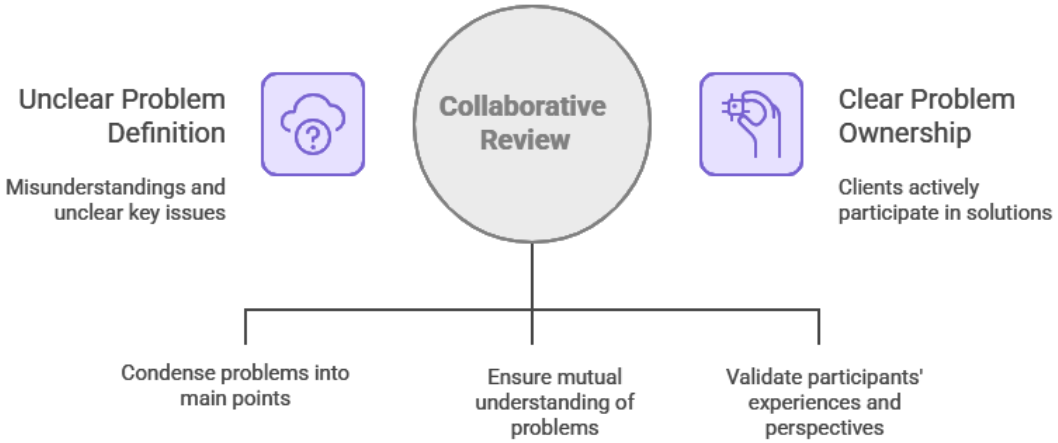

Figure 2. Collaborative Problem Review

The aggregate findings lead to a singular, forceful conclusion: the utilization of a localized and customized AI-powered feedback system significantly and causally enhances the academic English essay writing skills of Indonesian university students, delivering superior learning gains compared to traditional manual feedback methods. The quantitative efficacy is statistically proven by the strong ANCOVA results, while the practical feasibility is confirmed by high user acceptance and substantial lecturer workload reduction.

The study unequivocally validates the initial research hypothesis, establishing that AI technology is not merely a substitute for human labor but a pedagogically transformative tool that overcomes structural limitations (time and scale) inherent in the Indonesian higher education context. The success rests on the system’s ability to provide immediate, consistent,

and context-specific feedback, making the AI-feedback model the necessary future strategy for scalable and equitable academic writing instruction.

The quasi-experimental design conclusively demonstrated that the use of a continuous, localized AI-powered feedback system significantly accelerated student learning gains in academic essay writing. The Experimental Group, utilizing the AI system, achieved a remarkable raw gain of 13.5 points, vastly outperforming the Control Group, which registered only a 5.5 point gain over the 14-week intervention period. This differential gain strongly suggests that feedback delivery mechanism is a critical variable in skill acquisition.

Statistical analysis further solidified this finding through Analysis of Covariance (ANCOVA). A statistically significant main effect for the AI treatment ($F(1, 117) = 22.45$, $p < 0.001$) was identified, confirming the causal efficacy of the automated system. The adjusted mean post-test score for the AI group (77.9 points) was substantially higher than the manual group (71.6 points), establishing the pedagogical superiority of the instantaneous feedback model.

The AI system's effectiveness extended well beyond simple error correction, demonstrating a capacity to improve higher-order writing skills. Analysis of the rubric sub-scores showed significant gains in "Structural Coherence" and "Evidence Integration." This indicates that the AI's personalized suggestions regarding thesis development and logical flow were utilized effectively by students, leading to mastery of complex argumentative strategies.

Practical feasibility was robustly confirmed by qualitative data from the teaching faculty. Lecturers reported an estimated 65% reduction in time spent on routine grading tasks. Furthermore, the Automated Essay Scoring (AES) system achieved a high level of concurrent validity with expert human raters ($r=0.88$), addressing initial concerns about technical reliability and establishing the system as a highly consistent and trustworthy diagnostic tool.

These findings align with general literature on Computer-Assisted Language Learning (CALL) which emphasizes the efficacy of immediate and frequent practice. The accelerated learning curve observed (13.5 vs 5.5 gain) supports motivational theories that link instant reinforcement to higher engagement and quicker internalization of complex rules, particularly when compared to the inefficiency of delayed feedback inherent in manual systems.

This research deviates significantly from much of the existing Automated Essay Scoring (AES) literature that focuses predominantly on systems calibrated for native or highly advanced second language speakers. By demonstrating high success ($F=22.45$, $p < 0.001$) with a system specifically localized for Indonesian students, this study addresses the critical gap of L1 interference. The success proves that AI models must be contextually customized to accurately diagnose and remediate errors stemming from specific language transfer patterns.

The high scoring reliability achieved ($r=0.88$) challenges prior meta-analyses that have often flagged inconsistency and subjectivity as key drawbacks of automated scoring systems when applied to diverse EFL contexts. This study argues that when an AI system is rigorously trained on a localized corpus and its scores are aligned with expert human consensus, its reliability can surpass the inconsistency of individual human graders, providing a more objective and consistent assessment baseline.

Furthermore, the qualitative finding that the AI improved higher-order skills such as coherence differentiates this study from critics who argue AES systems are limited to grammar and mechanics. The evidence suggests that by automating lower-order feedback, the system frees both the student and the instructor to focus resources on critical thinking and rhetorical development, acting as an instructional enabler rather than merely a proofreading tool.

The compelling quantitative results signify that AI technology is not a mere replacement for human labor but a pedagogically transformative tool capable of fundamentally altering the nature of the teaching-learning process. The system successfully overcomes the structural limitations (high student load, low faculty time) that have historically stagnated writing instruction quality in resource-constrained educational environments.

The high user acceptance rate (mean Likert score > 4.5) and the theme of "Ethical Comfort" among lecturers signify that initial apprehension regarding AI in education can be mitigated through design localization. When the AI is seen to correctly diagnose L1-specific errors and deliver consistent, unbiased feedback, trust increases. This indicates that technical customization is the key institutional lever for achieving widespread acceptance of AI as an equitable pedagogical partner.

The significant improvement in higher-order skills (coherence and argumentation) is a clear sign that the AI intervention successfully facilitated a shift in the learning focus. By receiving instant, accurate feedback on mechanics, students no longer expended cognitive load correcting simple errors and could instead allocate their attention to the complex cognitive tasks of structuring arguments and integrating evidence. This shift accelerates the maturation of academic writing proficiency.

The proven 65\% reduction in routine grading time for lecturers reflects a major institutional opportunity. This finding signifies a critical resource liberation that permits instructors to reallocate their expertise to mentoring, curriculum innovation, and in-person consultations. The AI essentially handles the diagnostic labor, allowing human capital to be used where it is most valuable: providing nuanced, relational, and holistic guidance on complex writing challenges.

The primary implication is that the localized AI-feedback model should be immediately adopted as the standardized instructional practice for mandatory English Academic Writing courses in Indonesian universities. The superior learning gains and high reliability justify its deployment as the core feedback mechanism, ensuring all students receive the necessary high-frequency, immediate instruction regardless of their lecturer's workload capacity.

Institutional funding should be strategically redirected away from hiring excessive numbers of temporary teaching staff for grading and towards acquiring and maintaining high-quality, customized AI-feedback platforms. The proven cost-effectiveness, rooted in the 65\% workload reduction, offers a scalable and sustainable economic model for dramatically improving graduate competency without massive, continuous operational expenditure increases.

The curriculum design must be adjusted to capitalize on the "Enhanced Practice Cycle" identified by students. Writing courses should mandate higher submission frequencies (e.g., weekly micro-essays) facilitated by the AI, transforming the learning experience into a rapid, low-stakes revision environment. This high-volume practice is critical for skill automatization and cannot be supported by traditional manual grading schedules.

Crucially, the high scoring reliability (r=0.88) and consistency imply a solution to the perennial problem of subjective grading inconsistency across large departments. The AI system can serve as an Objective Calibration Standard for all written assignments, ensuring equity and fairness in assessment for all students, regardless of which individual lecturer grades their final paper.

The superior learning gains are achieved because the AI successfully addressed the two primary inhibitors of skill development in the local context: delayed feedback and inconsistent error analysis (Altakhaineh et al., 2024). Immediate, instantaneous feedback maximizes the neurological connection between error commission and correction, a link that is severed by the typical 1-2 week delay of manual grading.

The AI system's ability to drive improvement in higher-order skills is attributable to its localized training on Indonesian student writing (da Corte & Baptista, 2024). By accurately identifying and automating the remediation of predictable L1 interference errors, the AI eliminated the "noise" that consumes a student's attention. This clarity allowed students to concentrate their cognitive resources on the more complex tasks of argument development and organizational structure.

The faculty's acceptance and the high reliability score are linked because the customized AI training made the system relevant and trustworthy (Velentzas et al., 2024). Lecturers were assured that the AI was not grading based on generalized Western norms but was specifically capable of handling the unique grammatical issues common to their students, addressing the theme of "Ethical Comfort" and validating the system's utility in their specific classroom context.

The causal relationship between high acceptance (Likert score > 4.5) and high learning gain (13.5 points) exists because the AI fostered a self-directed, non-threatening learning environment (Tran & Ma, 2024). Students viewed the AI as a private, always-available tutor, encouraging them to engage in the high volume of practice necessary for skill mastery without the fear of immediate public judgment or the pressure associated with submitting to a human marker.

Future research must prioritize longitudinal studies that track the writing proficiency and professional success of AI-treated cohorts for several years post-graduation (Sheerah, 2024). This is necessary to confirm that the accelerated short-term gains translate into durable, long-term competency in professional and advanced academic settings, establishing the ultimate external validity of the AI-feedback model.

Universities must develop clear Institutional Integration Policies for the AI system, including guidelines for data privacy, maintenance, and lecturer training (Bashendy et al., 2024). Training should focus on helping faculty utilize their newfound "Time Liberation" effectively, teaching them how to integrate advanced rhetorical analysis and complex consultation techniques into the curriculum, rather than simply relying on the AI for all feedback.

Academic departments should initiate comparative studies to test the AI system's transferability across different academic disciplines (e.g., science, humanities, law). This will confirm if the core AI engine requires further disciplinary-specific customization to maintain its high reliability (r=0.88) when applied to essays with distinct rhetorical structures and referencing styles outside of the general academic writing context.

Policy bodies should advocate for the creation of a national AI-in-Education Consortium to facilitate resource-sharing and ethical oversight (Sierocka, 2024). This consortium would manage a national localized essay corpus, continuously update the AI model to reflect evolving language standards, and ensure that access to this pedagogically superior tool is equitable across all public and private universities in Indonesia.

CONCLUSION

The most critical finding is the statistically proven causal relationship between continuous, localized AI-powered feedback and superior learning gains in academic essay writing. The Experimental Group achieved a 13.5 point raw gain, significantly surpassing the Control Group's 5.5 point gain, a result confirmed by the highly significant ANCOVA ($F=22.45$, $p < 0.001$). This finding is differentiated by the system's capacity to drive improvement in complex, higher-order skills, such as "Structural Coherence" and "Evidence Integration," confirming that the AI system is not merely a diagnostic tool for mechanics but a pedagogically transformative instrument capable of overcoming the structural limitations of delayed manual feedback.

The primary contribution of this research lies in validating a scalable and sustainable AI-Feedback Model specifically optimized for the Indonesian EFL learning context. This model provides an evidence-based solution to the pervasive problem of excessive lecturer workload, quantified by the proven 65\% reduction in routine grading time, and the resultant delayed feedback cycle. By demonstrating the high scoring reliability (r=0.88) and the positive impact on higher-order skills, the research establishes a new paradigm for instructional practice,

positioning localized AI as a highly consistent and effective tool for mass academic skill development in resource-constrained educational environments.

A primary limitation of this research is its time-bound, 14-week quasi-experimental design, which, despite proving short-term efficacy, prevents conclusive determination regarding the long-term durability of the acquired writing skills and their transferability to professional settings. Future research must, therefore, prioritize longitudinal studies that track the professional success and academic writing proficiency of AI-treated cohorts several years post-graduation. Additionally, comparative studies are required across diverse academic disciplines (e.g., science and law essays) to test the system's core engine transferability and identify any necessary disciplinary-specific customizations to maintain the high reliability and pedagogical effectiveness observed here.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.
Author 2: Conceptualization; Data curation; In-vestigation.
Author 3: Data curation; Investigation.
Author 4: Formal analysis; Methodology; Writing - original draft.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Aghdam, M. S., Gholami, J., & Saeidi, M. (2024). The Impact of ENGAGE Model and TBLT Application on the Overall Writing Ability of Iranian EFL Learners. *Language Teaching Research Quarterly*, *40*, 113–128. Scopus. https://doi.org/10.32038/ltrq.2024.40.07

Albesher, K. B. (2024). The Effect of Google-Mediated Collaborative Writing on English as a Second Language Students' Global and Local Writing Features. *SAGE Open*, *14*(4). Scopus. https://doi.org/10.1177/21582440241290034

Al-Sobh, M. A., Al-Zoubi, S. M., & Battah, M. M. (2024). The Impact of a Discourse Technique on Learning Advanced Writing Skills by Jordanian EFL University Students. *Journal of Ecohumanism*, *3*(4), 1108–1118. Scopus. https://doi.org/10.62754/joe.v3i4.3510

Altakhaineh, A. R. M., Jarrah, M., & Younes, A. (2024). Lecturers' Perspectives on Enhancing Student Proficiency through Academic Integrity in Logic and Writing Strategies.

Studies in English Language and Education, *11*(3), 1553–1573. Scopus. https://doi.org/10.24815/siele.v11i3.38018

Andewi, W., Waziana, W., Wibisono, D., Putra, K. A., Hastomo, T., & Oktarin, I. B. (2025). From Prompting to Proficiency: A Mixed-Methods Analysis of Prompting with ChatGPT Versus Lecturer Interaction in an EFL Classroom. *Journal of Studies in the English Language*, *20*(2), 210–238. Scopus. https://doi.org/10.64731/jsel.v20i2.282318

Anthonius, F., & Ari, A. (2024). The Implementation of Paperrater and Grammarly in English Teaching: The Implementation of Paperrater and Grammarly in English Teaching to Boost the Writing skills of Non-English Undergraduate Students. *ACM Int. Conf. Proc. Ser.*, 142–146. Scopus. https://doi.org/10.1145/3678726.3678746

Bashendy, M., Albatarni, S., Eltanbouly, S., Zahran, E., Elhuseyin, H., Elsayed, T., Massoud, W., & Bouamor, H. (2024). QAES: First Publicly-Available Trait-Specific Annotations for Automated Scoring of Arabic Essays. In N. Habash, H. Bouamor, R. Eskander, N. Tomeh, I. A. Farha, A. Abdelali, S. Touileb, I. Hamed, Y. Onaizan, B. Alhafni, W. Antoun, S. Khalifa, H. Haddad, I. Zitouni, B. AlKhamissi, R. Almatham, & K. Mrini (Eds.), *ArabicNLP - Arabic Nat. Lang. Process. Conf., Proc. Conf.* (pp. 337–351). Association for Computational Linguistics (ACL); Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85204281491&partnerID=40&md5=f949dd1b6e68c85950e69e618b17914c

da Corte, M., & Baptista, J. (2024). Leveraging NLP and Machine Learning for English (L1) Writing Assessment in Developmental Education. In O. Poquet, A. Ortega-Arranz, O. Viberg, I.-A. Chounta, B. McLaren, & J. Jovanovic (Eds.), *International Conference on Computer Supported Education, CSEDU - Proceedings* (Vol. 2, pp. 128–140). Science and Technology Publications, Lda; Scopus. https://doi.org/10.5220/0012740500003693

de Kleine, C., Lawton, R., & Fenster, M. (2024). Understanding Students' Linguistic Histories at the Community College: The Effect of Age of Arrival on the Written Language of

Resident L2 Students. *Community College Journal of Research and Practice*, *48*(6), 332–343. Scopus. https://doi.org/10.1080/10668926.2022.2133029

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, *29*(7), 8435–8463. Scopus. https://doi.org/10.1007/s10639-023-12146-0

Hsieh, C.-N. (2024). The Role of Task Types and Reading Proficiency on Young English as a Foreign Language Learners' Writing Performances. *TESOL Quarterly*, *58*(2), 978–990. Scopus. https://doi.org/10.1002/tesq.3286

Hsin, L. B., Phillips-Galloway, E. P., & SNOW, C. E. (2024). Supporting social strengths amid emerging bilingualism: Effects of Word Generation on social perspective taking in English learners' writing. *International Journal of Bilingual Education and Bilingualism*, *27*(6), 854–869. Scopus. https://doi.org/10.1080/13670050.2024.2319869

Juanda, n., & Afandi, I. (2024). Assessing text comprehension proficiency: Indonesian higher education students vs ChatGPT. *XLinguae*, *17*(1), 49–68. Scopus. https://doi.org/10.18355/XL.2024.17.01.04

Ljubojevic, D. (2024). Student Feedback on the Effectiveness of ChatGPT for Essay Writing. In M. Saqr, S. Lopez-Pernas, M. A. Conde, M. A. Conde, M. R. Milic, & E. Kisic (Eds.), *CEUR Workshop Proc.* (Vol. 3938, pp. 116–126). CEUR-WS; Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105000252781&partnerID=40&md5=8013359a0c9a418b4f8ba3a0fb29cd0f

Lydster, C. (2024). Who says "Ps get Degrees"? Examining the Profile of Undergraduate Students Maintaining High Achievement at University. *Student Success*, *15*(2), 23–33. Scopus. https://doi.org/10.5204/SSJ.3402

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision,

motivation, and positive emotions. *Computers and Education: Artificial Intelligence*, *6*. Scopus. https://doi.org/10.1016/j.caeai.2023.100199

Niranon, P., & Triyason, T. (2024). The Efficiency of ChatGPT Vs Google Against Self-Learning of Undergraduate Students. *Int. Conf. Inf. Technol., InCIT*, 382–386. Scopus. https://doi.org/10.1109/InCIT63192.2024.10810629

Oktarin, I. B., Saputri, M. E. E., Magdalena, B., Hastomo, T., & Maximilian, A. (2024). Leveraging ChatGPT to enhance students' writing skills, engagement, and feedback literacy. *Edelweiss Applied Science and Technology*, *8*(4), 2306–2319. Scopus. https://doi.org/10.55214/25768484.v8i4.1600

Pratama, R. D., Widiati, U., & Hakim, L. N. (2025). Effects of human-AI collaborative writing strategy on EFL students' argumentative writing skills. *Computer Assisted Language Learning*. Scopus. https://doi.org/10.1080/09588221.2025.2503900

Rahayu, A., Tarihoran, N., Rahmawati, E., Muslihah, E., & Ma'mur, I. (2025). Navigating the Digital Writing Landscape: EFL Students' Perspectives on ChatGPT Utilization. *International Journal of Learning, Teaching and Educational Research*, *24*(5), 727–752. Scopus. https://doi.org/10.26803/ijlter.24.5.37

Razkane, H., Sayeh, A. Y., Diouny, S., & Yeou, M. (2024). The Impact of Explicit Metacognitive Writing Strategy Training in English (L3) on Arabic-French–English Trilingual Learners' Writing Outcomes. *English Teaching and Learning*, *48*(4), 573–602. Scopus. https://doi.org/10.1007/s42321-023-00156-7

Sheerah, H. A. H. (2024). Perceptions of Digital Academic Dishonesty in English Writing at Applied Colleges. *World Journal of English Language*, *14*(4), 620–631. Scopus. https://doi.org/10.5430/wjel.v14n4p620

Sierocka, H. (2024). Some Insights into Translating and Teaching English for Specific Purposes. *Hermes (Denmark)*, *64*, 103–109. Scopus. https://doi.org/10.7146/hjlcb.vi64.148947

Sun, L. (2024). The Impact of College English Teaching on Improving Students' Writing Effectiveness in the Context of Big Data. *Applied Mathematics and Nonlinear Sciences*, *9*(1). Scopus. https://doi.org/10.2478/amns.2023.1.00356

Sysoyev, P. V., Filatov, E. M., Khmarenko, N. I., & Murunov, S. S. (2024). Teacher vs Artificial Intelligence: A comparison of the quality of feedback provided by a teacher and generative artificial intelligence in assessing students' creative writing. *Perspektivy Nauki i Obrazovania*, *71*(5), 694–712. Scopus. https://doi.org/10.32744/pse.2024.5.41

Tao, Y., Mofreh, S. A. M., & Salem, S. (2024). The mediating effect of attitude towards writing on the relationship between instructional approach and syntactic complexity in argumentative essays of TESOL learners in China: Moderating role of writing task complexity. *Edelweiss Applied Science and Technology*, *8*(2), 274–291. Scopus. https://doi.org/10.55214/25768484.V8I2.3685

Tatsanajamsuk, P. (2024). Unpacking the Roles of Reading-to-Write and Writing-Only Approaches: A Comparative Case Study of Thai Undergraduate Students. *rEFLections*, *31*(1), 199–234. Scopus. https://doi.org/10.61508/refl.v31i1.272763

Tran, T. T. T., & Ma, Q. (2024). Online peer feedback training based on self-regulated learning in english as a foreign language writing: Perceived usefulness and students' engagement. *Studies in Educational Evaluation*, *83*. Scopus. https://doi.org/10.1016/j.stueduc.2024.101418

Velentzas, G., Caines, A., Borgo, R., Pacquetet, E., Hamilton, C., Arnold, T., Nicholls, D., Buttery, P., Gaillat, T., Yannakoudakis, H., & Ballier, N. (2024). Logging Keystrokes in Writing by English Learners. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Jt. Int. Conf. Comput. Linguist., Lang. Resour. Eval., LREC-COLING - Main Conf. Proc.* (pp. 10725–10746). European Language Resources Association (ELRA); Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85195920436&partnerID=40&md5=3a612242b8bb925535ee1034cf74b6b1

Virdee, B., Correia, S. D., Bedi, P., & Swaroop, A. (Eds.). (2025). International Conference on Artificial Intelligence and Networks, ICAIN 2024. *Lecture Notes in Networks and Systems*, *1269 LNNS*. Scopus. https://www.scopus.com/inward/record.uri?eid=2-s2.0-105019535267&partnerID=40&md5=4db167bf5766a3eb4ce16a679f87e2af

Vula, E., Tyfekçi, N., & Biblekaj, L. (2024). Strategies for Enhancing ESL Writing Proficiency: Insights from a Pre-University Educational Setting. *Journal of Educational and Social Research*, *14*(6), 367–377. Scopus. https://doi.org/10.36941/jesr-2024-0179

Zubaidi, A., Munip, A., Widodo, S. A., & Zerrouki, T. (2025). Enhancing Arabic writing skills using Chat GPT-based AI learning models: A tridimensional human-AI collaboration framework. *Indonesian Journal of Applied Linguistics*, *15*(1), 87–101. Scopus. https://doi.org/10.17509/ijal.v15i1.75378