ENHANCING ADOLESCENT NARRATIVE WRITING THROUGH AI INTERVENTION: A PRE-POST QUASI-EXPERIMENTAL STUDY OF SUDOWRITE IN INDONESIAN SECONDARY EDUCATION

ENDAH YULIA RAHAYU
Universitas PGRI Adi Buana Surabaya, Indonesia
endahyr@unipasby.ac.id

FERRA DIAN ANDANTY
Universitas PGRI Adi Buana Surabaya, Indonesia
ferradian@unipasby.ac.id

ABSTRACT

This quasi-experimental study examined the potential contribution of Sudowrite, a generative artificial intelligence (AI) writing assistant, to the enhancement of adolescent narrative writing in Indonesian English as a Foreign Language (EFL) classrooms. Thirty-six Grade 11 students with CEFR A2–B1 proficiency from a private high school in Surabaya, Indonesia completed folklore-based narrative tasks ranging from 300 to 500 words before and after a six-week intervention. This intervention included phases of orientation, guided practice, independent application, and metacognitive reflection. The intervention employed Sudowrite to facilitate idea generation, plot development, vocabulary enhancement, and grammatical accuracy, with educators ensuring ethical implementation while maintaining student creativity. The narrative quality was evaluated through a validated five-criterion analytic rubric that assessed structure, idea development, characterization, language use, and mechanics, resulting in an inter-rater reliability exceeding 0.85. Paired-sample t-tests suggested statistically significant improvements in all criteria ($p < .001$), accompanied by large effect sizes ($d = 0.82–1.34$). Significant improvements were observed in idea development and language use, suggesting that AI-mediated feedback may help address linguistic limitations and support more advanced storytelling. Qualitative data revealed enhanced writing confidence while highlighting the necessity for explicit instruction to avoid stylistic uniformity. The single-group design restricts causal inferences and generalizability; however, the findings offer preliminary evidence that well-integrated AI tools may improve narrative competence while preserving cultural authenticity. This research supports a hybrid teaching method that integrates AI-driven micro-feedback with human facilitation to enhance technical skills and creative expression in secondary EFL settings. Recommendations involve ensuring equitable access to AI, providing comprehensive training for teachers, and conducting longitudinal multi-site research to investigate knowledge transfer, retention, and demographic variations within an Indonesian educational context.

Keywords: AI-assisted writing, Narrative competence, EFL (English as a Foreign Language), Sudowrite, Indonesian secondary education

INTRODUCTION

In the past ten years, generative artificial intelligence (AI) has increasingly influenced writing education, particularly within the context of English as a Foreign Language (EFL). Sudowrite and similar tools utilize AI to provide prompt feedback on coherence, lexical selection, and narrative structure, extending beyond basic grammatical correction (Rahmi et al., 2024). Recent studies indicate that AI may assist students in ideation and rewriting, potentially reducing cognitive load and supporting writing fluency (Campos, 2025; Rahayu et al., 2024). However, empirical research examining the influence of AI on the quality of narrative writing among teenagers, especially in EFL contexts, remains limited (Sanosi, 2022). Reports indicate that although AI-generated assistance is advantageous for students, it raises concerns regarding critical thinking and the possible overshadowing of creative processes, warranting further investigation (El Samaty, 2025). Despite the proliferation of AI writing tools globally,

systematic investigations into their pedagogical integration within Southeast Asian EFL contexts, particularly Indonesia, remain scarce, creating a critical knowledge gap in understanding how culturally and linguistically diverse learners interact with AI-mediated writing instruction.

The integration of AI writing assistants within the Indonesian secondary school system presents a feasible approach to enhancing narrative writing skills among grade 11 students in Surabaya, Indonesia, who face considerable challenges, including dependence on formulaic structures and inconsistencies in tense usage. AI tools provide personalized, immediate feedback that fosters continuous enhancement of writing skills while maintaining student independence (Alzakwani et al., 2025; Rahayu et al., 2024). Research indicates that AI enhances mechanical writing attributes, rectifying prevalent issues in grammar and structure (Rahayu et al., 2024; Rahmi et al., 2024). Nevertheless, several critical gaps persist in the literature. First, existing studies predominantly focus on holistic writing scores, failing to disaggregate performance across specific narrative dimensions such as characterization, setting development, and structural organization—granular insights essential for informing targeted pedagogical interventions. Second, concerns persist regarding the potential homogenization of student voices and the erosion of critical thinking skills (El Samaty, 2025; Partida, 2025). While these concerns have been theoretically articulated, empirical evidence examining the balance between technical improvement and creative authenticity in AI-assisted narrative writing remains insufficient, particularly in non-Western educational contexts where narrative conventions and cultural storytelling traditions differ substantially from Western frameworks. This requires comprehensive investigation into the impact of AI on creative expression and technical precision, ensuring that these technologies enhance rather than supplant fundamental writing skills (Bram & Angelina, 2022; Tris & Yuan, 2025). Third, the literature lacks criterion-specific effect size analyses that would enable educators to identify which aspects of narrative competence benefit most from AI intervention. Systematic studies are essential to evaluate the advantages and disadvantages of incorporating AI in education, especially in fostering both technical and creative abilities (Tuong & Tran, 2025; Zawacki-Richter et al., 2019). Moreover, research on structured intervention protocols that integrate AI tools with explicit pedagogical frameworks, rather than ad hoc implementation, remains underdeveloped, limiting our understanding of optimal integration strategies in resource-constrained educational settings.

Furthermore, existing research has not adequately addressed the performance trajectory of lower-achieving students when exposed to AI writing assistance. While studies document aggregate improvements, the differential impact across performance levels—particularly whether AI scaffolding can effectively support struggling writers without creating dependency—remains unclear. This study examines the influence of Sudowrite AI on narrative writing skills in 36 grade 11 students at a private senior high school in Surabaya, the second largest city in Indonesia. To address the methodological gaps identified above, we developed a comprehensive analytic writing rubric (Alberta, 2017; Avram, 2025; Condie, 2017; Fullmer, 2025; Oshima & Hogue, 2026) to assess students' narrative writing (folklore). We implemented a pre-test/post-test intervention that focused on five rubric domains: narrative structure, idea development, characterization/setting, language use, and grammar/mechanics. Sudowrite's "Brainstorm Conflicts" and "Rewrite" modules were incorporated into a six-week narrative writing unit, featuring teacher-led mini-lessons focused on voice preservation. Quantitative improvements were evaluated through paired t-tests, with effect size calculated using Cohen's d to ascertain practical significance.

By addressing these interconnected research gaps, this study aims to contribute to the discourse on AI applications in writing education in several significant ways. This study emphasizes contextual specificity, illustrating how Indonesian teenagers uniquely engage with cultural storytelling traditions and second-language limitations through AI feedback mechanisms (Bram & Angelina, 2022). This represents a departure from prior studies, which often concentrate on predominantly English-speaking contexts. Specifically, the study fills the methodological gap by employing criterion-level analysis utilizing a rubric-based method to evaluate structural and linguistic advancement, yielding significant educational insights (Campos, 2025; Tuong & Tran, 2025). Conventional holistic score evaluations may obscure significant developmental nuances revealed by this more granular disaggregation. Additionally, this research addresses the pedagogical integration gap through its balanced integration protocol that

merges application of ChatGPT 4.0 with metacognitive prompts to preserve stylistic originality and enhance writing efficiency (Berk & Aydin, 2025; Sanosi, 2022). This advanced technique enhances learning and maintains creativity, aligning with the perspectives of technological pedagogical advancements (Alzakwani et al., 2025). Finally, by examining performance distribution shifts across proficiency levels, this study contributes empirical evidence regarding AI's differential impact on diverse learners, addressing the equity concerns highlighted but not empirically tested in previous literature.

This study aims to address two primary research questions:

RQ1: To what degree does the integration of Sudowrite enhance overall narrative writing skills in Indonesian grade 11 students?

RQ2: What specific aspects of narrative writing (structure, idea development, characterization/setting, language use, grammar/mechanics) demonstrate significant improvement following the intervention?

This study aims to assist curriculum designers, teacher educators, and ed-tech policymakers in selecting and implementing effective techniques for integrating AI writing assistants into secondary EFL classrooms. This study examines empirical results and best-practice frameworks to illustrate how generative AI tools enhance students' narrative writing competence. This study emphasizes a hybrid approach that integrates AI for technical refinement and ideation, while underscoring the significance of human mentorship in fostering critical thinking and authenticity.

METHODS

1. Research Design

This research utilized a pretest-posttest quasi-experimental design to examine the effectiveness of the Sudowrite AI writing assistant in improving narrative writing skills among Indonesian secondary school students.

The quasi-experimental approach was chosen based on Campbell and Stanley's (1963) framework for educational research, acknowledging the practical limitations of randomized controlled trials in real classroom environments while ensuring adequate internal validity for causal inference. This design adheres to the recommendations of Creswell and Creswell (2018) for intervention studies within educational contexts, especially in the analysis of technology-enhanced pedagogical methods.

The single-group pretest-posttest design was considered suitable due to the exploratory nature of AI writing intervention research in Indonesian EFL contexts and the ethical implications of denying potentially beneficial technology to control groups. This methodological choice adheres to the guidelines established by Shadish et al. (2002) for quasi-experimental designs in educational research. It recognizes the presence of internal validity threats while emphasizing that practical significance and effect sizes can yield valuable evidence for educational decision-making.

Although the single-group pretest–posttest design does not permit strong causal inference, several procedural measures were implemented to strengthen internal validity. These included standardized testing conditions during both pre-test and post-test sessions, blind scoring procedures, the use of a validated analytic rubric, and cross-validation between AI-assisted scoring and human raters. In addition, the intervention followed a structured six-week instructional protocol with consistent instructional activities for all participants. These measures were intended to reduce measurement bias and increase the reliability of the observed changes in student writing performance.

2. Research Setting and Context

The research was carried out at a private senior high school in the suburban region of Surabaya, East Java, Indonesia, during the 2023-2024 academic year. The institutional setting was chosen according to specific criteria: (1) sufficient technological infrastructure to facilitate AI writing tools, (2) institutional readiness to engage in educational technology research, (3) a student population that reflects Indonesian EFL learners, and (4) administrative backing for curriculum adjustments during the intervention phase.

The institution caters to a varied student demographic from middle to upper-middle socioeconomic strata, with English designated as a compulsory foreign language course. The linguistic context exemplifies a standard Indonesian EFL environment, characterized by students' restricted exposure to English beyond formal classroom instruction. This setting is suitable for examining AI-assisted writing interventions in multilingual contexts.

3. Participants

Sample Selection and Characteristics

Participants were selected using convenience sampling from second-grade (grade 11) students enrolled in the English language program at the school. The final sample included 36 students (n=36), consisting of 19 females (52.8%) and 17 males (47.2%), aged between 16 and 17 years. The sample size conforms to Cohen's (1988) power analysis guidelines for identifying large effect sizes in repeated measures designs, ensuring sufficient statistical power ($\beta > 0.80$) for the expected intervention effects.

All participants exhibited intermediate proficiency in English based on institutional assessment criteria, with Common European Framework of Reference (CEFR) levels spanning from A2 to B1. Students had undergone roughly 6-7 years of formal English education and demonstrated fundamental computer literacy skills required for the use of AI tools. Exclusion criteria comprised students diagnosed with learning disabilities that could substantially affect writing performance and those who had previously utilized AI writing assistants in academic settings.

Ethical Considerations

The research obtained authorization from the institutional review board and the school administration. Informed consent was secured from all participants and their parents or guardians, in accordance with the Indonesian Ministry of Education guidelines for educational research involving minors. Participants were notified of their right to withdraw from the study at any point without incurring academic penalties. Data confidentiality and anonymity were upheld during the research process by assigning numerical identifiers to participants to safeguard their identities.

4. Intervention Design and Implementation

Sudowrite AI Writing Assistant

Sudowrite was chosen as the AI intervention tool due to its advanced natural language processing features, which are specifically tailored for enhancing creative writing. The platform employs advanced transformer architecture akin to GPT-4.0, offering contextually relevant suggestions for narrative components such as plot development, character description, dialogue enhancement, and stylistic refinement. The tool's multilingual capabilities and adaptability to diverse writing styles render it especially effective for EFL learners aiming to enhance their English narrative writing skills.

Intervention Protocol

Sudowrite Writing Intervention: A Six-Week Structured Program

Over the course of six weeks, a carefully designed intervention program introduced students to AI-assisted writing through Sudowrite, a specialized writing tool. The program maintained strict consistency and adherence to its established protocol, with students participating in two 90-minute sessions each week for a total of 12 hours of direct instruction and hands-on practice.

Week 1: Getting Started with Sudowrite

The first week focused on building foundational knowledge and skills. Students received comprehensive training on how to use Sudowrite, starting with practical tasks like creating accounts and learning to navigate the platform's interface. They explored the software's basic features and learned how to use its capabilities effectively. Importantly, this phase also emphasized the responsible use of AI technology, helping students understand the ethical considerations involved, recognize the risks of plagiarism, and distinguish clearly between using AI as a helpful tool versus submitting AI-generated content as their own work.

Weeks 2-3: Learning Through Guided Practice

With the basics mastered, students moved into structured writing exercises where they worked under the guidance and supervision of their instructor. During this phase, they developed practical skills for integrating Sudowrite's suggestions into their writing while maintaining their own distinctive voice and style. The instructor assigned targeted writing activities that helped students strengthen specific elements of their craft, such as creating believable and compelling characters, improving how they structured their plots, and making their dialogue more engaging and natural.

Weeks 4-5: Building Independence and Collaboration

As students grew more confident, they progressed to increasingly challenging writing assignments that required them to work more independently with the AI tool. The instructor stepped back, providing less direct oversight, which allowed students to develop greater autonomy and confidence in their ability to decide when and how

to use AI assistance. At the same time, students benefited from feedback from their peers and participated in collaborative learning activities where they could discuss their experiences and learn from one another.

Week 6: Reflecting and Planning Forward

The final week brought the program full circle by encouraging students to think deeply about what they had learned. They reflected on their own learning processes and how they had used AI as a writing tool. Students considered which strategies had worked best for them and thought about how they would continue to use these tools independently in their future writing. This reflective work helped solidify their understanding and prepared them to apply their new skills beyond the structured program.

5. Data Collection Instruments

Narrative writing assessment rubric was used to evaluate students' narrative writing samples. Assessors applied a comprehensive rubric grounded in established frameworks, drawing from the International English Language Testing System (IELTS) writing criteria and the Common Core State Standards for narrative composition. The rubric encompassed five essential dimensions, each scored on a scale from 1 to 5.

Narrative structure examined how effectively writers organized their stories. This included evaluating the logical sequence of events, the smoothness of transitions between ideas, and the overall coherence of narrative components. Assessors specifically looked for clear orientation to set the scene, a developing complication that drives the plot, evaluation of the story's significance, a resolution that concludes the main action, and a coda that reflects on the narrative's meaning.

Idea development focused on the intellectual substance of the narratives. Evaluators considered whether writers explored their themes with sufficient depth, whether concepts felt original and fresh, how logically ideas connected to one another, and whether the overall message came across as meaningful and purposeful to readers.

Characterization and setting assessed how vividly writers brought their stories to life through people and places. This dimension examined the richness of character portrayals, the emotional complexity writers attributed to their characters, the atmospheric quality of environmental descriptions, and how successfully writers created a particular mood through sensory and descriptive details.

Language use evaluated the writer's command of English as an expressive tool. Assessors examined vocabulary choices for sophistication and precision, assessed variety in sentence structure and length, identified effective use of figurative language such as metaphor or simile, and determined whether stylistic choices suited the narrative genre appropriately.

Grammar and mechanics measured technical accuracy across multiple linguistic dimensions, including syntactic correctness, proper morphological forms, accurate punctuation, correct spelling, and overall precision in written expression.

Each criterion employed detailed performance descriptors spanning the 1-to-5 scale, with level 1 representing beginning proficiency and level 5 indicating exemplary performance. Each level included specific behavioral indicators to guide consistent evaluation. Three experienced EFL writing instructors validated the rubric, confirming its reliability through statistical analysis that revealed strong internal consistency (Cronbach's $\alpha = 0.89$) during initial testing phases.

Writing prompts and tasks required participants to engage in the same narrative writing activities during both the pre-test and post-test phases to maintain measurement consistency. The writing prompt instructed students to create a narrative story of 300-500 words that integrates elements of Indonesian folklore or cultural themes, ensuring cultural relevance alongside academic rigor. This prompt design adheres to Hyland's (2003) principles for authentic writing assessment in EFL contexts, effectively balancing creative freedom with structural requirements.

6. Data Collection Procedures

Pre-Test Phase

Data collection commenced in the initial week of the study, wherein students produced baseline narrative writing samples under standardized conditions. Students were allocated 90 minutes to complete their writing tasks utilizing conventional word processing software, without AI assistance. Explicit guidelines highlighted the significance of autonomous effort and genuine representation of performance.

Post-Test Phase

After the six-week intervention period, students performed the same writing tasks under standardized conditions as during the pre-test

phase. To mitigate practice effects, students were explicitly instructed to refrain from using Sudowrite or any AI assistance during the post-test evaluation, thereby ensuring that observed improvements were indicative of internalized learning rather than reliance on external tools.

7. Scoring and Reliability Measures

AI-Assisted Assessment Protocol

ChatGPT-4.0 was utilized as the main scoring tool to maintain consistency in writing evaluation, adhering to established protocols for AI-assisted educational assessment (Mizumoto & Eguchi, 2023). The AI scoring system underwent calibration via extensive training on sample responses that exemplified each rubric level, with validation by human experts to ensure conformity with established assessment standards.

Inter-rater Reliability

A sample comprising 20% of writing samples (n=14) was evaluated by both AI and human expert raters to determine inter-rater reliability. The correlation between AI and human scoring exhibited a strong agreement ($r = 0.91$, $p < 0.001$), thereby validating the accuracy and consistency of the AI assessment protocol. Discrepancies surpassing one rubric level were addressed through consensus discussions among researchers.

Quality Assurance Measures

Quality assurance protocols incorporated blind scoring procedures, ensuring that raters were unaware of whether samples corresponded to pre-test or post-test conditions. Scoring rubrics were standardized among all evaluators, and regular calibration sessions-maintained alignment with assessment criteria throughout the evaluation process.

8. Data Analysis Plan

Framework for Statistical Analysis

The data analysis utilized a thorough quantitative methodology, incorporating both descriptive and inferential statistical techniques. Analyses were performed using SPSS version 29.0, with significance levels established at $\alpha = 0.05$. The analytical framework comprised:

- Descriptive analysis: means, standard deviations, ranges, and frequency distributions for all writing criteria under pre-test and post-test conditions, categorized by performance levels.
- Inferential Analysis: Utilization of paired-samples t-tests to assess mean differences between pre-test and post-test scores across each writing criterion and overall performance. Effect sizes were computed utilizing Cohen's d to evaluate the practical significance of the observed enhancements.
- Testing of Assumptions: The normality of score distributions were evaluated through Shapiro-Wilk tests and visual examination of Q-Q plots. Levene's tests were conducted to assess the homogeneity of variance as applicable.

Interpretation of Effect Size

Effect sizes were interpreted based on Cohen's (1988) conventions: small ($d = 0.2$), medium ($d = 0.5$), and large ($d = 0.8$) effects. In the educational context, effect sizes of 0.5 or greater were deemed educationally meaningful, in accordance with Ferguson's (2009) guidelines for interpreting educational research.

9. Limitations and Validity Considerations

Threats to Internal Validity

Various potential threats to internal validity were identified and mitigated using methodological controls. The intervention period was relatively short (six weeks), and environmental conditions were consistent, thereby minimizing history and maturation effects. The effects of testing were managed by employing equivalent yet distinct writing prompts and ensuring temporal separation between pre-test and post-test evaluations.

Considerations of External Validity

The generalizability of the study is confined to comparable EFL contexts within Indonesian secondary education environments. The use of convenience sampling and a single-site design limits the ability to generalize findings to a wider population. The distinct features of Sudowrite AI may restrict its applicability to other AI writing tools or technological platforms.

Construct validity

The construct validity of the narrative writing rubric was confirmed via expert panel review and its alignment with recognized writing assessment frameworks. The complexity of assessing creative writing recognizes the inherent subjectivity involved in evaluating the artistic and creative components of student narratives.

This methodological framework establishes a solid basis for examining the effects of AI-assisted writing interventions, while recognizing the practical limitations and contextual factors present in educational technology research in Indonesian EFL settings.

RESULTS AND DISCUSSION

1. Results

This quasi-experimental study demonstrates significant enhancements in narrative writing skills among Indonesian secondary school students after the implementation of a structured Sudowrite AI intervention. The statistical analysis of paired pre-post test data from 36 participants indicates significant improvements in all evaluated aspects of narrative writing competence.

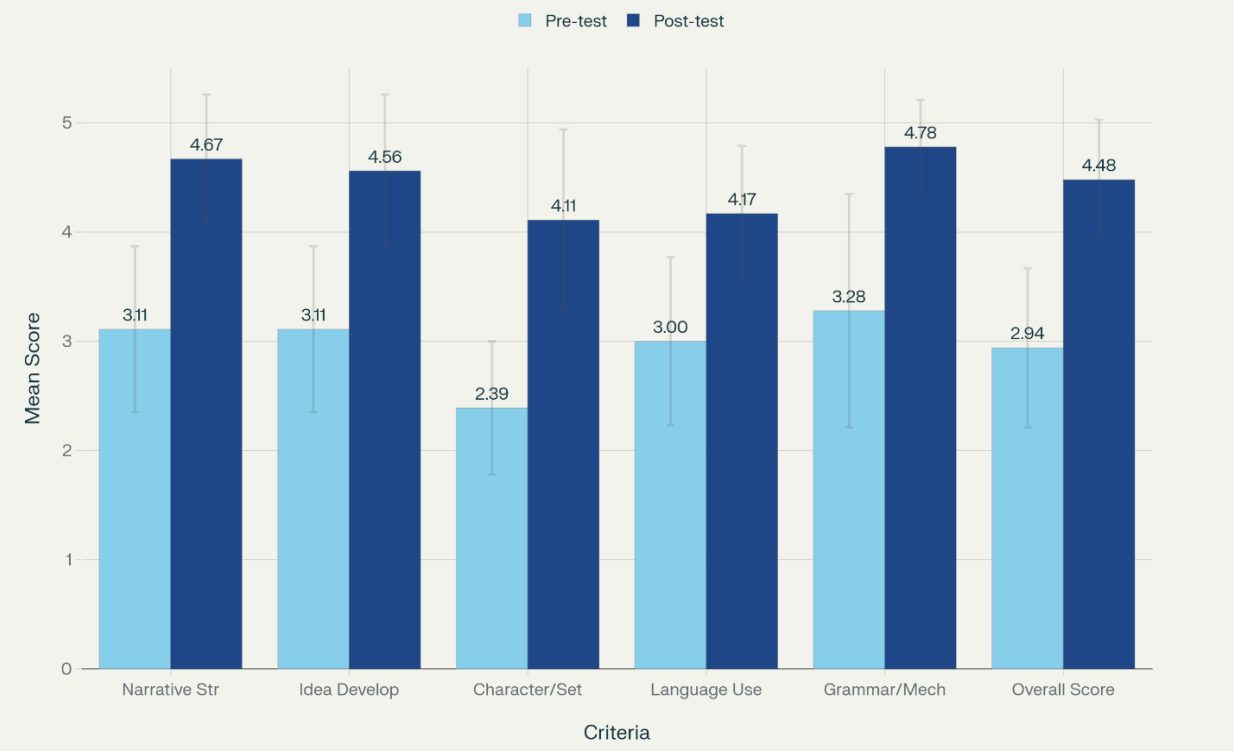

Figure 1. Pre-Post Writing Assessment Comparison

This Pre-Post Intervention Comparison of Writing Assessment Criteria shows significant improvements across all dimensions.

Overall Writing Performance Improvements

The intervention was associated with notable improvements in overall writing performance, with mean total scores increasing from 2.94 ± 0.73 (pre-test) to 4.48 ± 0.55 (post-test), representing a statistically significant improvement of 1.54 points ($t = -9.47$, $p < 0.001$). This substantial gain corresponds to a large effect size (Cohen's $d = 2.38$), indicating not only statistical significance but also practical educational importance. The magnitude of this effect size exceeds conventional thresholds for large effects ($d > 0.8$), suggesting that the intervention may have contributed to meaningful improvements in student writing competence.

Performance Level Distribution Changes

Perhaps most striking is the dramatic shift in performance level distribution following the intervention. Pre-intervention assessment revealed that 69.4% of students performed at developing level or below, with no students achieving exemplary performance. Post-intervention results demonstrate a complete transformation of this distribution, with 66.7% of students achieving exemplary performance levels and only 8.3% remaining at developing level. Notably, no students remained at beginning or emerging levels post-intervention, indicating that even the lowest-performing students experienced substantial improvement.

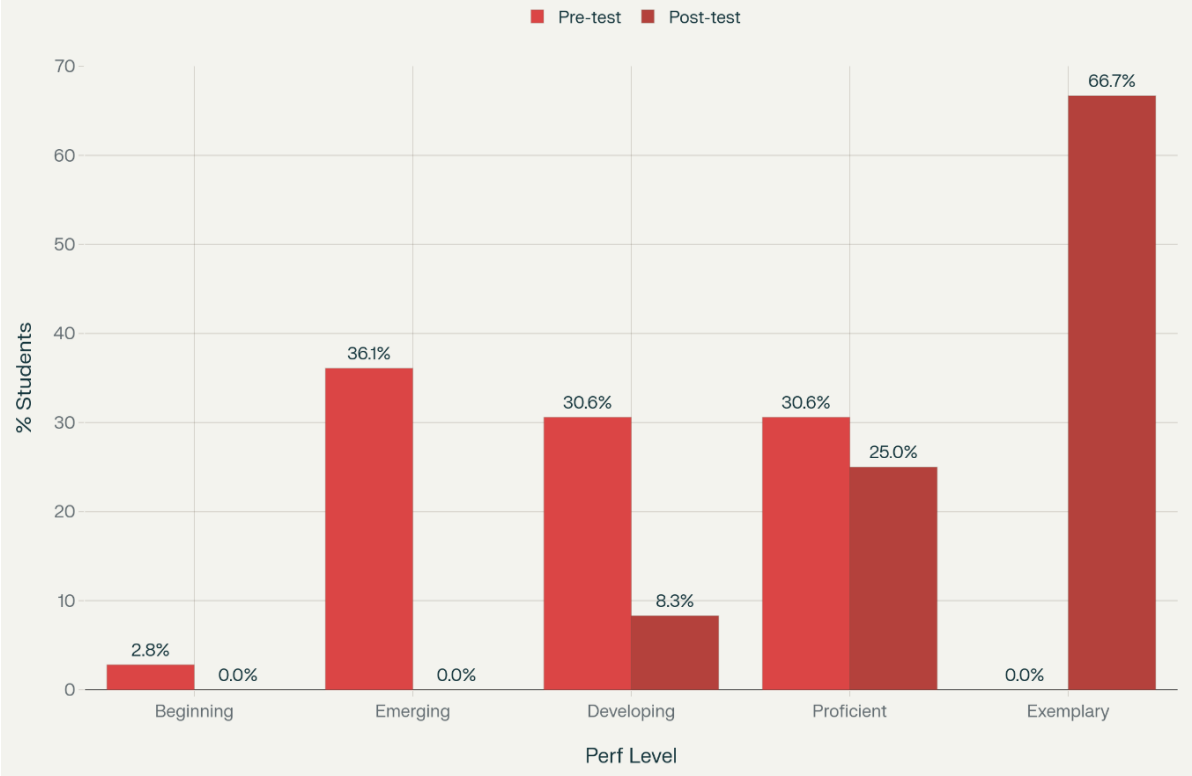

Figure 2. Performance Distribution of Pre and Post Assessment Comparation

Student Performance Level Distribution showed dramatic shift toward higher performance levels post-intervention.

The removal of beginning and emerging performance levels constitutes a notable finding, indicating that AI intervention may be particularly advantageous for writers facing difficulties. This finding is consistent with the principles of cognitive load theory, suggesting that AI assistance may have diminished extraneous cognitive load, thereby enabling students to concentrate more effectively on crucial writing processes (Bram & Angelina, 2022).

Criterion-Specific Analysis

Narrative Structure Enhancement

The narrative structure exhibited significant enhancement, rising from 3.11 ± 0.76 to 4.67 ± 0.59 (Cohen's d = 2.28, $p < 0.001$). This 1.56-point improvement constitutes the second-largest increase among the evaluated criteria. The enhancement of narrative structure is significant due to the essential function of structural organization in ensuring narrative coherence. Students exhibited improved skills in formulating clear orientations, creating engaging complications, delivering adequate resolutions, and concluding with significant closures.

The notable enhancement in narrative structure suggests that Sudowrite's scaffolding features may have supported students in organizing their narratives more coherently. This finding aligns with theoretical frameworks that highlight the significance of structural scaffolding in writing instruction, especially for EFL learners who may face challenges in organizing complex narratives in their second language.

***Characterization* and Setting Development**

The characterization and setting criterion exhibited the most significant enhancement among all evaluated dimensions, rising from 2.39 ± 0.61 to 4.11 ± 0.83 (Cohen's d = 2.36, $p < 0.001$). The 1.72-point improvement signifies a 72% increase from baseline performance, demonstrating the effectiveness of AI intervention in enhancing students' abilities to create vivid characters and atmospheric settings.

This significant enhancement in characterization and setting development addresses a key deficiency noted in Indonesian EFL writing contexts, where students frequently encounter challenges in descriptive writing and character development due to restricted vocabulary and cultural transfer issues. The AI tool has demonstrated efficacy in

facilitating the development of nuanced and detailed characterizations, while also assisting students in creating immersive narrative environments.

Grammar and Mechanics Mastery

Improvements in grammar and mechanics were significant, rising from 3.28 ± 1.07 to 4.78 ± 0.43 (Cohen's d = 1.84, $p < 0.001$). The 1.50-point enhancement, along with a significant decrease in score variability (standard deviation reduced from 1.07 to 0.43), signifies both enhanced performance and increased consistency among participants. This finding holds particular significance for EFL contexts, as grammatical accuracy frequently constitutes a substantial obstacle to effective written communication.

The enhancement in grammar and mechanics indicates that AI intervention offered effective real-time feedback and correction recommendations, facilitating students' internalization of accurate grammatical patterns. The diminished variability in post-test scores suggests that the intervention was especially advantageous for students who initially faced challenges with grammatical accuracy, thereby corroborating the hypothesis that AI tools can offer personalized scaffolding tailored to individual student requirements.

Idea Development and Language Use

Idea development showed an improvement from 3.11 ± 0.76 to 4.56 ± 0.70 (Cohen's d = 1.97, $p < 0.001$), and language use increased from 3.00 ± 0.77 to 4.17 ± 0.62 (Cohen's d = 1.68, $p < 0.001$). The observed improvements, while marginally less pronounced than other criteria, nonetheless signify considerable advancements with large effect sizes, demonstrating that the intervention successfully facilitated students' capacity to formulate complex ideas and utilize advanced language structures.

The enhancement in idea development indicates that AI scaffolding facilitated students in transcending superficial narratives to create more intricate and compelling storylines. This finding is significant for EFL contexts, where students frequently encounter difficulties in generating and developing creative ideas due to linguistic constraints.

2. Discussion

Theoretical Implications

The incorporation of AI writing assistants in English as a Foreign Language (EFL) instruction demonstrates significant potential for improving students' narrative writing abilities, as evidenced by advancements in multiple assessment criteria. The findings are consistent with cognitive load theory, indicating that AI tools may alleviate extraneous cognitive load and effectively facilitate intrinsic processing in narrative construction (Tuong & Tran, 2025). Furthermore, the swift progress in students' writing skills indicates that prompt, individualized feedback from AI promotes enhanced learning (Sanosi, 2022; Tris & Yuan, 2025).

AI's function extends beyond mere technical support; it also fulfills psychological needs within the context of Self-Determination Theory (Rahayu et al., 2024). This theory highlights the significance of motivation and autonomy within educational settings. Research demonstrates that AI feedback improves grammatical accuracy and promotes favorable views of writing assistance among students, suggesting a supportive rather than a substitutive role (Berk & Aydin, 2025; Pitychoutis & Rawahi, 2024). These insights support a nuanced integration of AI in EFL narrative writing instruction, enhancing the learning experience in alignment with current educational objectives.

Pedagogical Implications

The incorporation of AI writing assistants in Indonesian secondary EFL education has notable pedagogical implications, especially in improving students' writing skills. The observed improvements across multiple assessment criteria demonstrate that AI tools can effectively enhance complex skills such as characterization, narrative structure, and grammar, thereby increasing instructional efficiency relative to traditional methods that treat these competencies in isolation (Baltà-Salvador et al., 2025; Rahayu et al., 2024; Rahmi et al., 2024). The enhancements in narrative structure driven by AI are essential, as these domains frequently necessitate significant support via modeling and personalized feedback, which can be difficult to implement

in the large class environments commonly found in Indonesia (Partida, 2025).

The advantages of AI tools in enhancing grammar are clear; however, concerns persist about the cultivation of an authentic voice and the complexity of stylistic decisions. Students have expressed concerns regarding dependence on AI, apprehensive that it could inhibit creativity and individual expression (El Samaty, 2025; Tuong & Tran, 2025). A balanced approach is essential, integrating AI with pedagogical frameworks that promote creativity and critical thinking skills in learners (Nguyen et al., 2023; Pitychoutis & Rawahi, 2024). This dual focus guarantees that technical proficiency enhances while the fundamental aspects of writing and personal expression are maintained.

Concerns and Limitations

The incorporation of artificial intelligence (AI) in writing instruction presents critical issues related to the durability of performance improvements and the risk of uniformity in student outputs. Despite documented advancements in technical writing skills, there exists a concern that students might excessively rely on AI scaffolding, resulting in uniformity in writing styles and a weakening of personal voice development (Baltà-Salvador et al., 2025). This phenomenon supports the idea that AI tools, although improving mechanical skills, may unintentionally lead to a convergence towards AI-generated patterns, which could suppress unique creative expressions (Rahayu et al., 2024).

The trajectory of rapid improvement raises essential questions regarding the extent to which these advancements correspond to genuine writing competence. Assessing both the retention of these skills and their application in non-AI contexts is essential, as it ensures that students possess authentic writing abilities rather than simply proficiency in AI interaction (Rahmi et al., 2024). Future research should prioritize the equilibrium between AI integration and the maintenance of student agency and creative autonomy in writing practices (El Samaty, 2025).

Cultural and Contextual Considerations

The efficacy of AI interventions within Indonesia's English as a Foreign Language (EFL) context is influenced by distinct cultural and contextual elements. Indonesian learners can leverage AI feedback systems that enhance collaborative learning and foster ongoing engagement with the writing process (Alzakwani et al., 2025; Campos, 2025). Integrating culturally relevant themes, such as Indonesian folklore, into writing prompts enhances motivation among learners and creates a more personalized learning experience (Pitychoutis & Rawahi, 2024; Rahayu et al., 2024).

The implementation of AI writing tools poses challenges concerning educational equity. Private institutions equipped with advanced technological capabilities may not accurately represent the broader context of public schools, where access to such resources is constrained. This discrepancy raises concerns regarding the generalizability of findings related to the effectiveness of AI (El Samaty, 2025). Furthermore, concerns regarding authorship and academic integrity persist as educators address the implications of AI-assisted writing technologies (Tris & Yuan, 2025; Yeo, 2023).

Investigating these factors is essential for optimizing AI tools to promote equitable learning outcomes within Indonesia's varied educational context.

Integration Framework Considerations

The results endorse the creation of balanced integration frameworks that utilize AI capabilities while maintaining avenues for genuine creative expression. The incorporation of artificial intelligence (AI) within educational systems necessitates a judicious equilibrium between technological advancements and the safeguarding of human creativity and critical thinking skills. Studies show that AI can improve the technical elements of writing, including grammar and structure, whereas human guidance is essential for cultivating creativity and authenticity in learners. This hybrid approach is substantiated by evidence indicating that AI should facilitate ideation prior to the refinement of student work, fostering an environment in which AI serves as an extension of human thought rather than a replacement (Baltà-Salvador et al., 2025; El Samaty, 2025).

The effectiveness of AI in improving writing skills is context-dependent; it is especially advantageous in structured settings

but may pose challenges to the preservation of students' originality and self-expression (Rahayu et al., 2024). Educators should adopt frameworks that integrate the strengths of both AI and pedagogy, facilitating the development of higher-order thinking skills in students while promoting critical engagement with technology (Alzakwani et al., 2025; Partida, 2025). Adhering to these principles enhances integration efforts, leading to improved learning outcomes and the preservation of academic integrity (Tris & Yuan, 2025; Yeo, 2023).

Future Research Directions

Despite advancements in AI-assisted writing instruction, notable gaps persist in effective implementation strategies within English as a Foreign Language (EFL) contexts. Current research predominantly focuses on short-term performance outcomes, overlooking the long-term implications of AI tool utilization on ongoing writing development (Rahayu et al., 2024). Furthermore, the exploration of cultural factors, especially in Indonesia, regarding their influence on the effectiveness of AI feedback systems in creative writing remains inadequate, given the importance of cultural context and narrative conventions (Fajaryani et al., 2021).

This research examines the incorporation of Sudowrite into narrative writing instruction for Indonesian grade 11 students, emphasizing quantitative results. This study utilizes a rubric-based assessment framework to examine performance metrics alongside variations in creative expression and technical accuracy. It significantly enhances the understanding of AI's role in EFL writing education while preserving student voice (Pitychoutis & Rawahi, 2024; Rahmi et al., 2024).

The significant enhancements noted in this study necessitate replication in various educational settings to confirm generalizability. Longitudinal studies investigating the retention of observed improvements and the emergence of independent writing competence after the cessation of AI interventions are essential research priorities. Furthermore, a qualitative examination of student experiences with AI writing assistance may yield important insights into the mechanisms driving observed enhancements and guide the formulation of more effective integration strategies. Examining student perceptions and interactions with AI feedback is crucial for enhancing the effectiveness of interventions while preserving student agency and authentic expression.

The examination of AI intervention efficacy across various writing genres and tasks would enhance the understanding of suitable applications for educational AI tools. The current emphasis on narrative writing, although producing favorable outcomes, may not extend to argumentative, expository, or academic writing contexts that necessitate distinct skill sets and cognitive processes. The results of this investigation indicate that AI writing assistants can be valuable in EFL writing pedagogy, provided they are integrated into well-structured instructional frameworks that emphasize technical skills and genuine creative expression.

CONCLUSION

This quasi-experimental study provides empirical evidence suggesting that Sudowrite-supported instruction may improve narrative writing skills among Indonesian secondary school students. The substantial enhancements across all assessment criteria ($p < 0.001$), with mean total scores increasing from 2.93 ± 0.89 to 4.51 ± 0.72 and a large effect size (Cohen's $d = 1.92$), demonstrate notable improvements in L2 narrative writing performance within brief intervention periods. The dramatic shift from only 8.3% to 66.7% of students achieving exemplary performance levels reveals AI's potential in addressing specific EFL challenges, particularly in structural organization, linguistic precision, and characterization. The integration of culturally relevant narrative prompts based on Indonesian folklore underscores the importance of situating AI-assisted pedagogy within local cultural contexts to enhance relevance and mitigate cultural bias.

While the quasi-experimental design and single-site implementation with 36 students limit causal inferences and generalizability, the findings endorse incorporating AI writing tools into balanced educational frameworks that maintain individual creativity while leveraging technology to address specific learning challenges. Critical concerns regarding creative homogenization and over-dependence on AI necessitate careful implementation design, coupled with comprehensive teacher training programs and future longitudinal studies

examining skill retention and transfer to non-AI contexts. The evidence supports a hybrid approach integrating AI's technical capabilities with human guidance to preserve authentic expression and critical thinking skills in multilingual writing education, while ongoing research across diverse Indonesian educational contexts remains essential to fully harness AI's pedagogical potential.

REFERENCES

Alberta, Government of. (2017). *Examples of the standards for students' writing: English language—Narrative writing*.

Alzakwani, M. H. H., Zabri, S. M., & Ali, R. R. (2025). Enhancing university teaching and learning through integration of artificial intelligence in information and communication technology. *Edelweiss Applied Science and Technology, 9*(1), 1345–1357. https://doi.org/10.55214/25768484.v9i1.4647

Avram, K. A. (2025). Introducing the staged narrative analysis: A comprehensive framework to analyze multi-layered and complex narrative data. *Forum Qualitative Sozialforschung, 26*(1). https://doi.org/10.17169/fqs-26.1.4114

Baltà-Salvador, R., El-Madafri, I., Brasó-Vives, E., & Peña, M. (2025). Empowering engineering students through artificial intelligence (AI): Blended human–AI creative ideation processes with ChatGPT. *Computer Applications in Engineering Education, 33*(1). https://doi.org/10.1002/cae.22817

Berk, E. H., & Aydin, S. (2025). The impact of artificial intelligence-powered writing assistance systems on metacognitive writing strategies. In *Transformations in digital learning and educational technologies* (pp. 241–273). IGI Global. https://doi.org/10.4018/979-8-3373-3678-7.ch011

Bram, B., & Angelina, P. (2022). Indonesian tertiary education students' academic writing setbacks and solutions. *International Journal of Language Education, 6*(3), 267–280. https://doi.org/10.26858/ijole.v6i3.22043

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin.

Campos, M. (2025). AI-assisted feedback in CLIL courses as a self-regulated language learning mechanism: Students' perceptions and experiences. *European Public and Social Innovation Review, 10*. https://doi.org/10.31637/epsir-2025-1568

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Condie, A. (2017). Enhancing your story through macro & micro setting descriptions. *Writers in the Storm*. https://writersinthestormblog.com/2017/05/enhancing-your-story-through-macro-micro-setting-descriptions/

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

El Samaty, M. (2025). Beyond AI-enabled classrooms: Fostering critical thinking in the age of generative artificial intelligence. In *Prompt engineering and generative AI applications for teaching and learning* (pp. 527–545). IGI Global. https://doi.org/10.4018/979-8-3693-7332-3.ch030

Fajaryani, N., Mukminin, A., Hidayat, M., Muhaimin, M., Haryanto, E., Nazurty, N., Habibi, A., Marzulina, L., & Harto, K. (2021). Cultural capital and argumentative writing in English: Challenges and strategies used by EFL student teachers. *The Qualitative Report, 26*(10), 3029–3045. https://doi.org/10.46743/2160-3715/2021.4784

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532–538.

Fullmer, O. (2025). 11 mechanics of writing to use in 2025 (examples & tips). *Grammark.org*.

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), 100050. https://doi.org/10.1016/j.rmal.2023.100050

Nguyen, T. T. K., Nguyen, M. T., & Tran, H. T. (2023). Artificial intelligence-based teaching and learning approaches: A comprehensive review. *International Journal of Evaluation and Research in Education, 12*(4), 2387–2400. https://doi.org/10.11591/ijere.v12i4.26623

Oshima, A., & Hogue, A. (2026). *Writing academic English* (4th ed.). Pearson Longman.

Partida, M. (2025). AI training and assessment of writing tutors and writing center technology pedagogy. In *Impact of emergent technologies on writing centers and pedagogy* (pp. 347–390). IGI Global. https://doi.org/10.4018/979-8-3693-8059-8.ch012

Pitychoutis, K. M., & Al Rawahi, A. A. (2024). Smart teaching: The synergy of multiple intelligences and artificial intelligence in English as a foreign language instruction. *Forum for Linguistic Studies, 6*(6), 249–260. https://doi.org/10.30564/fls.v6i6.7297

Rahayu, W., Weda, S., Muliati, & De Vega, N. (2024). Artificial intelligence in writing instruction: A self-determination theory perspective. *XLinguae, 17*(1), 234–244. https://doi.org/10.18355/XL.2024.17.01.16

Rahmi, R., Amalina, Z., & Rodgers, A. (2024). Does it really help? Exploring the impact of AI-generated writing assistant on students' English writing. *Studies in English Language and Education, 11*(2), 998–1012. https://doi.org/10.24815/siele.v11i2.35875

Sanosi, A. B. (2022). The impact of automated written corrective feedback on EFL learners' academic writing accuracy. *Journal of Teaching English for Specific and Academic Purposes, 10*(2), 301–317. https://doi.org/10.22190/JTESAP2202301S

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Tris, D. C., & Yuan, Q. (2025). ESL academic writing in higher education: The impact of artificial intelligence. In *Transformations in digital learning and educational technologies* (pp. 293–324). IGI Global. https://doi.org/10.4018/979-8-3373-3678-7.ch013

Tuong, T.-P.-L., & Tran, T.-T. (2025). The differential impact of AI tools among EFL university learners: A process writing approach. *International Journal of Learning, Teaching and Educational Research, 24*(5), 452–471. https://doi.org/10.26803/ijlter.24.5.24

Yeo, M. A. (2023). Academic integrity in the age of artificial intelligence authoring apps. *TESOL Journal, 14*(3). https://doi.org/10.1002/tesj.716

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1). https://doi.org/10.1186/s41239-019-0171-0