International Journal of Research in Education

Volume 6, Issue 2, July 2026, pp. 300 – 317

e-ISSN: 2745-3553

DOI: https://doi.org/10.26877/ijre.v6i2.3902

Reducing Dependence on AI-Generated Text through Contextual Writing Assessment

Yeni Daniarti[1*], Yanuardi[2]

[1,2]*Universitas Muhammadiyah Tangerang*

*Corresponding author's email: yenidaniarti@umt.ac.id

ARTICLE INFO

Received: June 1, 2026

Revised: July 5, 2026

Accepted: July 16, 2026

Keywords:

AI-generated text, contextual writing assessment, academic integrity, learner agency.

ABSTRACT

Although generative artificial intelligence (GAI) increasingly impacts academic writing and integrity in higher education, empirical evidence on assessment strategies to reduce students' reliance on AI-generated texts remains limited. This study aimed to examine the effectiveness of the Contextual Writing Assessment in reducing AI-generated text dependence and improving writing performance, writing originality, academic integrity, and learner agency. A quasi-experimental mixed-methods design with an embedded explanatory approach was employed in an English course at the Faculty of Engineering at the University of Muhammadiyah Tangerang. The participants comprised 150 second-semester students from six classes and four lecturers. Three classes were assigned to the experimental group and received Contextual Writing Assessment, while three classes served as the control group and received the conventional writing assessment. Data were collected through writing tests, AI-dependence questionnaires, academic integrity and learner agency questionnaires, originality scores, AI-use logs, reflective journals, draft analyses, observations, and interviews. The results demonstrated the effectiveness of the intervention, as the experimental group outperformed the control group in reducing AI-generated text dependence and exhibited substantial improvements in writing performance, originality, academic integrity, and learner agency. Qualitative evidence indicated more transparent AI use, meaningful revisions, stronger reflections, and reduced copy-paste behavior. The study concludes that Contextual Writing Assessment offers an ethical, process-based, and AI-resilient approach to writing assessment.

Introduction

Generative artificial intelligence has rapidly transformed higher education by providing students and lecturers with tools for generating, revising, translating, summarizing, and evaluating academic texts (Susilo et al., 2026). In English writing instruction, AI-based tools such as ChatGPT, Gemini, and similar platforms offer new opportunities for brainstorming, vocabulary enrichment, grammar checking, and feedback. However, these tools also create pedagogical and ethical challenges. Students may use AI not as a learning aid but as a substitute for thinking, drafting, revising, and constructing original ideas (Hapsari & Sudarman, 2025; Rafida et al., 2024). This raises a critical concern in higher education: how

can lecturers maintain writing authenticity, academic integrity, and students' ownership of learning in the age of AI-generated text?

Recent literature indicates that AI integration in higher education should not be understood merely as technology adoption. It must be examined as a pedagogical transformation that involves learning design, assessment validity, ethical governance and student agency. (Imran et al., 2024) argued that AI tools can enhance learning systems through personalized learning, adaptive tutoring, digital learning environments, and AI-assisted assessment; however, their implementation must address ethical issues, equity, and institutional readiness. Similarly, (Abolina et al., 2026) found that students and academic staff recognized the potential of AI-enriched tools to improve the study process, but they also identified challenges related to academic integrity, competence gaps, and the need for responsible AI use. These findings imply that AI use in higher education requires structured pedagogical control rather than unrestricted technological freedom(Fauziah et al., 2025; Febriyanti et al., 2026; Yunita et al., 2025).

The problem becomes more serious in writing assessments. Traditional writing assignments often emphasize final products such as essays, reports, or paragraphs (Septyaningrum et al., 2024). In the context of generative AI, final-product assessment is increasingly vulnerable because students can submit AI-generated texts that appear coherent, grammatically accurate, and academically acceptable. (Gonsalves, 2025) emphasized that generative AI challenges the reliability and authenticity of conventional assessments because it can generate text, audio, visual content, reflective writing, reports, case studies, and other academic products. Therefore, assessment in the age of AI must evolve from merely evaluating the final text to evaluating contextual understanding, writing process, ethical decision-making, and authentic student contribution.

Contextual assessment offers a relevant response to these challenges. Contextual assessment requires students to connect writing tasks to real-life, disciplinary, professional, experiential, situational, and ethical contexts. In English courses for engineering students, contextual writing assessments may require students to produce technical descriptions, process explanations, problem-solution paragraphs, short reports, and reflective writing on engineering topics. This approach makes writing tasks more resistant to direct AI substitution because students must demonstrate personal reasoning, contextual evidence, revision history, and responsibility for the final product (Hidayat et al., 2025; Lumintang & Kuswandono, 2025; Rohadi et al., 2025). (Gonsalves, 2025) proposed that assessment design in the age of generative AI should incorporate practical, situational, experiential, interdisciplinary, and ethical dimensions to strengthen higher-order thinking and real-world application.

The literature on AI adoption also suggests that AI tools become educationally meaningful only when they fit academic tasks and are supported by clear guidelines. (Soodan et al., 2024) showed that task-technology fit and perceived usefulness influence academics' continued use of AI chatbots, while also highlighting the importance of transparency regarding plagiarism, privacy, and ethical implications. This finding is important for English writing instruction because students' use of AI cannot be managed only through prohibition. Instead, students need explicit academic rules, AI-use documentation, reflective tasks, and

lecturer feedback to ensure that AI supports learning, rather than replacing it. Thus, contextual writing assessments should not reject AI entirely but should regulate its use in the writing process (Faridah et al., 2026; Firdes, 2024; Gunawan & Susanto, 2025).

In addition, writing in engineering education requires more than just grammatical accuracy. It requires the ability to organize technical ideas, explain processes, solve problems, and communicate disciplinary knowledge. (Khuder & Negretti, 2025) argued that writing is a crucial but often overlooked component of active, collaborative, and project-based learning in STEM contexts. Their study showed that effective collaborative writing depends on regulation, shared goals, metacognitive engagement, and instructor support. This insight is relevant to the present study because students' dependence on AI-generated text may indicate weak writing regulation, limited self-monitoring, and insufficient learner agency. Therefore, writing assessments should include drafting, peer or lecturer feedback, revision, and reflection (Asrifan et al., 2025; Pariyanto & Tungka, 2025; Sulistyanto et al., 2025; Ulfah, 2026).

Although previous studies have discussed AI adoption, AI-enriched learning, contextual assessment, and writing regulation, several gaps remain in the literature. First, many studies focus on the usefulness and adoption of AI tools, but fewer studies have examined how assessment design can reduce students' dependence on AI-generated text. Second, contextual assessment has been proposed as a theoretical response to generative AI, but empirical validation across disciplines remains limited. Third, research on AI in higher education often discusses opportunities and ethical risks, but it has not sufficiently examined how AI-use logs, reflective journals, draft analyses, and contextual writing rubrics can be integrated into English writing assessments. Fourth, although studies in STEM and engineering education have highlighted the importance of writing regulation, few have investigated it in the specific context of English courses for engineering students in Indonesian higher education.

Based on these gaps, this study proposes the Contextual Writing Assessment as a pedagogical and assessment strategy to reduce students' dependence on AI-generated text. The novelty of this study lies in the integration of contextual writing tasks, ethical AI guidance, AI-use logs, reflective journals, draft analysis, originality assessment, and lecturer judgment. Instead of treating AI use as misconduct by default, this study positions AI as a controlled support tool. Students are allowed to use AI for limited purposes, such as brainstorming, grammar checking, vocabulary support, and clarity improvement, but they must document, revise, and justify their use of AI. This makes the assessment process more transparent, ethical, and learning-oriented.

This study was conducted in an English course for second-semester students at the Faculty of Engineering, University of Muhammadiyah Tangerang. The participants consisted of 150 students from six classes and four lecturers. This study employed a quasi-experimental mixed-methods design with an embedded explanatory approach. Three classes were assigned to the experimental group and received the Contextual Writing Assessment, while three other classes were assigned to the control group and received the conventional writing assessment. Quantitative data were collected through writing tests, AI-dependence questionnaires, academic integrity questionnaires, learner agency

questionnaires, originality scores, and perception questionnaires. Qualitative data were collected through AI-use logs, reflective journals, writing drafts, observations, and semi-structured interviews.

The research problems of this study are formulated as follows: (1) What is the level of students' dependence on AI-generated text before the implementation of the Contextual Writing Assessment? (2) Does Contextual Writing Assessment significantly reduce students' dependence on AI-generated text? (3) Does Contextual Writing Assessment improve students' writing performance, writing originality, academic integrity, and learner agency? (4) How do students and lecturers perceive the practicality, relevance, and usefulness of Contextual Writing Assessment in English courses for engineering students?

In line with these research problems, the objectives of this study are to identify students' initial dependence on AI-generated text, examine the effectiveness of Contextual Writing Assessment in reducing AI-generated text dependence, measure its impact on writing performance, writing originality, academic integrity, and learner agency, and explore students' and lecturers' perceptions of its implementation. The study proposes the following hypotheses: H1: Contextual Writing Assessment significantly reduces students' dependence on AI-generated text. H2: Students who receive Contextual Writing Assessment demonstrate higher writing originality than those who receive conventional writing assessments. H3: Contextual Writing Assessment significantly improves students' English writing performance. H4: Contextual writing assessments positively influence students' academic integrity and learner agency. H5: Students and lecturers show positive perceptions of the implementation of Contextual Writing Assessment.

This study is significant because it addresses a current problem in higher education: the increasing dependence on AI-generated text in academic writing. The study contributes to the literature by extending the discussion of contextual assessment, AI ethics, learner agency, and writing regulation in the context of AI-supported English writing. It contributes practically by offering a classroom-based assessment model that can help lecturers evaluate not only the final writing product but also the process of thinking, drafting, revising, documenting AI use, and reflecting on their academic responsibility. Therefore, this study is expected to provide an empirical basis for designing more ethical, contextual, and AI-resilient writing assessments in higher education.

Research Methods

This study employed a quasi-experimental mixed-methods design with an embedded explanatory approach (Daniarti et al., 2026). This design was selected because the study was conducted in naturally existing classroom settings, where individual random assignment was not practically feasible. The quantitative strand was used to examine the effectiveness of the Contextual Writing Assessment in reducing students' dependence on AI-generated text and improving their English writing performance, writing originality, academic integrity, and learner agency. The qualitative strand was embedded to explain the quantitative findings by exploring students' writing processes, AI-use behavior, reflective awareness, and lecturers' perceptions of the assessment model.

The quantitative design used a non-equivalent control group pre-test and a post-test design. Six existing English classes were included in the study. Three classes were assigned to the experimental group and received Contextual Writing Assessment, while the other three classes were assigned to the control group and received the conventional writing assessment. The research design is presented in Table 1.

Table 1 Quasi-experimental Design

Group Type	Design	Condition
Experimental	O1 – X – O2	Contextual Writing Assessment
Control	O1 – C – O2	Conventional Writing Assessment

Note. 01 = pre-test; 02 = post-test; X = Contextual Writing Assessment; C = Conventional Writing Assessment

O1 refers to the pre-test, which included a writing test, AI-dependence, academic integrity, and learner agency questionnaires. X refers to the implementation of Contextual Writing Assessment. C refers to the conventional writing assessment. O2 refers to the post-test, which included a writing test, AI-dependence questionnaire, academic integrity questionnaire, learner agency questionnaire, writing product analysis, AI-use log, reflective journal, observation and interviews. This design enabled the researcher to compare students' learning outcomes before and after the intervention and examine differences between students who experienced Contextual Writing Assessment and those who experienced conventional writing assessment.

This study was conducted at the Faculty of Engineering, Universitas Muhammadiyah Tangerang, Indonesia. The research context was an English course for second-semester engineering students. English is a compulsory course that supports students' academic communication, technical writing, and professional-language competence. The writing tasks were developed based on engineering-related contexts, including technical descriptions, process explanations, short reports, problem-solution paragraphs, reflective paragraphs, and engineering-related academic writing. These tasks were designed to encourage students to write based on their own reasoning, disciplinary context, classroom discussions, and revision processes, rather than relying heavily on AI-generated text.

This study did not prohibit the use of AI tools. Instead, AI has been positioned as a controlled learning support tool. Students in the experimental group were guided to use AI ethically for limited purposes, such as brainstorming, vocabulary support, grammar checking, sentence improvement, and feedback on clarity of writing. However, they were not allowed to submit AI-generated texts as final products without meaningful revisions, reflections, and documentation.

The population of this study consisted of second-semester students from the Faculty of Engineering, Universitas Muhammadiyah Tangerang, who were enrolled in the English course. The sample consisted of 150 students from six intact classes and four English lecturers. The experimental group consisted of approximately 75 students from three classes, and the control group consisted of approximately 75 students from three classes. The four lecturers were involved as teaching practitioners, observers, evaluators and interview informants. The sampling technique used in this study was cluster sampling, as the participants were selected based on existing class groups. Before treatment, both groups

were tested to examine their initial equivalence in writing performance, AI-generated text dependence, academic integrity, and learner agency.

The independent variable of this study was Contextual Writing Assessment. It is a writing assessment model that requires students to produce English texts based on academic, technical, professional, and real-world engineering contexts. The assessment emphasized contextual relevance, personal reasoning, drafting, revision, transparency in AI use, reflective writing, and academic responsibility. The main dependent variable was students' dependence on AI-generated text, which refers to the extent to which students rely on AI tools to generate ideas, write paragraphs, produce final drafts, and complete writing assignments without sufficient personal contribution and revision. The supporting dependent variables were English writing performance, writing originality, academic integrity, learner agency, and students' and lecturers' perceptions of Contextual Writing Assessment.

Several instruments were used to collect data. First, contextual writing tasks were developed to connect English writing with engineering-related academic and professional contexts in the course. Second, a writing performance rubric was used to assess students' writing products based on content relevance, organization, vocabulary, grammar, mechanics, coherence, contextual appropriateness, and technical relevance of their writing. The writing scores ranged from 0 to 100. Third, an AI-generated text dependence questionnaire was used to measure students' dependence on AI-generated text using a five-point Likert scale. The questionnaire covered the frequency of AI use, reliance on AI for idea generation, reliance on AI for paragraph writing, reliance on AI for final draft production, lack of revision after AI use, difficulty writing without AI, and perceived need for AI in writing tasks.

Fourth, an academic integrity questionnaire was used to measure honesty, transparency, responsibility, ethical AI use, awareness of plagiarism, and writing ownership. Fifth, a learner agency questionnaire was used to measure goal-setting, self-regulation, decision-making, self-revision, confidence, and responsibility. Sixth, an originality rubric and similarity-checking procedure were used to evaluate students' writing originality. Seventh, an AI-use log was used to document the AI tool used, purpose of use, prompts, AI output summary, accepted parts, rejected parts, revisions made, reasons for revision, and ethical statement. Eighth, a reflective journal was used to explore students' awareness of their writing processes. Ninth, an observation checklist was used to monitor the implementation of Contextual Writing Assessment. Tenth, semi-structured interview guides were used to collect deeper information from the selected students and lecturers.

The treatment was implemented in several stages. First, students in both groups completed a writing pre-test and questionnaires on AI dependence, integrity, and learner agency. Second, the students in the experimental group received an orientation on ethical AI use in academic writing. Third, the lecturer introduced contextual writing topics related to engineering. Fourth, students generated ideas independently by preparing outlines, key vocabulary, and initial arguments before using AI. Fifth, students were allowed to use AI for limited academic support, but each use had to be documented in an AI-use log. Sixth, the students wrote their first drafts and received feedback from their peers and lecturers.

Seventh, the students revised their drafts and wrote reflective journals explaining their writing process, AI-use decisions, revision strategies, and efforts to maintain originality. Eighth, students submitted the final writing product together with the initial draft, revised draft, AI-use log, reflective journal and ethical statement. Ninth, students completed the post-test and questionnaires, and selected students and lecturers were interviewed to clarify quantitative findings.

Data were collected using quantitative and qualitative techniques. Quantitative data were obtained from the writing pre- and post-tests, AI-dependence questionnaire, academic integrity questionnaire, learner agency questionnaire, originality rubric, similarity score, perception questionnaire, and treatment fidelity checklist. Qualitative data were obtained from AI-use logs, reflective journals, writing drafts, observation notes, semi-structured interviews, peer and lecturer feedback records and students' ethical statements. Quantitative data were used to determine whether the Contextual Writing Assessment was effective, while qualitative data were used to explain how and why the intervention influenced students' writing behavior and AI-use responsibility.

The validity of the instruments was examined through expert judgment. Experts in English language teaching, educational assessment, and educational technology reviewed the writing rubric, questionnaires, AI-use log, reflective journal guide, observation checklist, and interview guide. The reliability of the questionnaires was tested using Cronbach's Alpha, with a minimum acceptable coefficient of 0.70. The reliability of the writing assessment was examined through inter-rater reliability by involving at least two raters who used the same analytical rubric. The consistency of the intervention was examined through treatment fidelity using an observation checklist with a minimum implementation criterion of 80%. The trustworthiness of the qualitative data was ensured through triangulation, member checking, and audit trails. Triangulation was conducted by comparing data from AI-use logs, reflective journals, draft writing, interviews, observation notes, and writing products.

Quantitative data were analyzed using descriptive and inferential statistical analyses. Descriptive statistics were used to calculate the mean, standard deviation, minimum score, maximum score, and percentage. Normality and homogeneity tests were conducted before hypothesis testing. An independent-samples t-test was used to examine the initial equivalence of the experimental and control groups. A paired-sample t-test was used to compare pre- and post-test scores within each group, while an independent-sample t-test was used to compare post-test and gain scores between the experimental and control groups. N-gain was calculated to measure improvement, and Cohen's d was used to determine the effect size. Correlation analysis was used to examine the relationships among AI-generated text dependence, writing performance, writing originality, academic integrity, and learner agency. Regression analysis was used to examine whether academic integrity and learner agency predicted a reduction in AI-generated text dependence.

Qualitative data were analyzed using thematic analyses. The analysis involved reading the data repeatedly, coding meaningful units, grouping codes into categories, identifying major themes, interpreting the findings, and connecting themes with quantitative results. The main themes included responsible AI use, reduced copy-paste behavior, improved self-revision, contextual awareness, ethical reflection, writing confidence, and controlled AI

assistance. The final interpretation was conducted by integrating the quantitative and qualitative findings to provide both statistical evidence and contextual explanations.

This study followed the ethical principles of educational research. Students were informed of the research objectives, procedures, data use, confidentiality, and their right to participate voluntarily. Students' identities were kept confidential, and the data were used only for academic research purposes. The study did not penalize students for using AI. Instead, it encourages ethical, transparent, and responsible AI use. Each student in the experimental group was required to submit an ethical statement explaining how AI was used and to confirm personal responsibility for the final writing product.

Findings

This section presents the findings on the effectiveness of Contextual Writing Assessment in reducing students' dependence on AI-generated text. The findings are organized into quantitative and qualitative results sections. The quantitative results compared the experimental and control groups in terms of AI-generated text dependence, English writing performance, writing originality, academic integrity, learner agency, and perceptions. The qualitative results explain students' AI-use behavior, reflective awareness, writing development, and lecturers' responses to the interventions.

Quantitative Findings

Table 2 presents the descriptive changes and between-group hypothesis tests for the principal quantitative outcomes of the study. It compares the experimental group, which received Contextual Writing Assessment, with the control group, which received conventional writing assessment. The analysis covers students' dependence on AI-generated text, writing originality, similarity scores, writing performance, N-gain, academic integrity, and learner agency. Because the two groups demonstrated closely comparable pre-test means, the subsequent differences in gain or change scores provide a reasonable basis for evaluating the relative effectiveness of the intervention. Welch's independent-samples *t*-tests were used to compare the groups, while Cohen's *d* and Hedges' *g* were reported to indicate the magnitude of the observed differences.

Table 2 Descriptive Changes dan Between-Group Hypothesis Tests

Outcome	Experimental group	Control group	MD (SE)	*t (df)*	Exact *p*	95% CI	Effect size (*d* / *g*)	Decision
AI-generated text dependence	3.76 -> 2.48 ΔM = -1.28 (2.75)	3.71 -> 3.36 ΔM = -0.35 (1.82)	-0.930 (0.381)	-2.442 (128.390)	.01595	[-1.683, -0.177]	-0.399 / -0.397	H1 supported
Writing originality	64.30 -> 80.40 ΔM = 16.10 (30.60)	64.00 -> 70.20 ΔM = 6.20 (20.70)	9.900 (4.266)	2.321 (130.000)	.02186	[1.460, 18.340]	0.379 / 0.377	H2 supported

Outcome	Experimental group	Control group	MD (SE)	*t (df)*	Exact *p*	95% CI	Effect size (*d* / *g*)	Decision
Similarity score	34.80 -> 18.60 ΔM = -16.20 (33.15)	34.20 -> 28.70 ΔM = -5.50 (22.45)	-10.700 (4.623)	-2.315 (130.081)	.02221	[-19.846, -1.554]	-0.378 / -0.376	H2 supported
Writing performance	62.40 -> 78.90 ΔM = 16.50 (24.70)	62.10 -> 70.60 ΔM = 8.50 (16.70)	8.000 (3.443)	2.324 (129.961)	.02170	[1.189, 14.811]	0.379 / 0.378	H3 supported
N-gain	M = 0.44 SD = 0.65	M = 0.22 SD = 0.43	0.220 (0.090)	2.445 (128.359)	.01586	[0.042, 0.398]	0.399 / 0.397	H3 supported
Academic integrity	3.04 -> 4.08 ΔM = 1.04 (2.09)	3.02 -> 3.38 ΔM = 0.36 (1.41)	0.680 (0.291)	2.336 (129.801)	.02103	[0.104, 1.256]	0.381 / 0.380	H4 supported
Learner agency	3.08 -> 4.02 ΔM = 0.94 (1.83)	3.06 -> 3.41 ΔM = 0.35 (1.24)	0.590 (0.255)	2.311 (130.121)	.02238	[0.085, 1.095]	0.377 / 0.376	H4 supported

Note. *N = 150 (n = 75 per group). Values in parentheses after ΔM are standard deviations. Welch's independent-samples t-tests compared individual gain or change scores; exact two-tailed p values are reported. MD = experimental group minus control group. Negative MD values indicate favorable reductions for AI-generated text dependence and similarity scores. d = Cohen's d; g = Hedges' g.*

The clearest change concerned students' dependence on AI-generated text. In the experimental group, the mean score decreased from 3.76 before the intervention to 2.48 after the intervention. By comparison, the control group showed only a modest decrease, from 3.71 to 3.36. The difference between the two groups was statistically significant, $MD = -0.93$, $SE = 0.38$, $t(128.39) = -2.44$, $p = .01595$, 95% CI [−1.68, −0.18]. Although the effect was small to moderate, Hedges' $g = -0.40$, the finding remains educationally important. It indicates that students who experienced Contextual Writing Assessment were less likely to rely on AI as a substitute for their own thinking and writing.

Students in the experimental group also demonstrated stronger growth in writing originality. Their mean originality score increased from 64.30 to 80.40, while the control group increased from 64.00 to 70.20. The difference in improvement was statistically significant, $MD = 9.90$, $SE = 4.27$, $t(130.00) = 2.32$, $p = .02186$, 95% CI [1.46, 18.34], with Hedges' $g = 0.38$. This improvement was supported by the similarity-score results. The experimental group's similarity score decreased from 34.80% to 18.60%, whereas the control group's score decreased from 34.20% to 28.70%. The greater reduction in the experimental group was significant, $MD = -10.70$, $SE = 4.62$, $t(130.08) = -2.31$, $p = .02221$, 95% CI [−19.85, −1.55], Hedges' $g = -0.38$. Together, these findings suggest that students were not merely changing words or lowering similarity mechanically. They were increasingly producing texts that reflected their own reasoning, contextual understanding, and revision decisions.

Writing performance showed a similar pattern. The experimental group improved from a mean score of 62.40 to 78.90, producing a mean gain of 16.50. The control group improved from 62.10 to 70.60, with a smaller mean gain of 8.50. The difference between the groups

was statistically significant, *MD* = 8.00, *SE* = 3.44, *t*(129.96) = 2.32, *p* = .02170, 95% CI [1.19, 14.81], Hedges' *g* = 0.38. The N-gain result further confirms this pattern. The experimental group achieved an average N-gain of 0.44, compared with 0.22 in the control group. This difference was significant, *MD* = 0.22, *SE* = 0.09, *t*(128.36) = 2.45, *p* = .01586, 95% CI [0.04, 0.40], Hedges' *g* = 0.40. These results indicate that Contextual Writing Assessment supported more meaningful improvement in students' English writing abilities.

The intervention also influenced students' ethical awareness and ownership of learning. Academic integrity in the experimental group increased from 3.04 to 4.08, while the control group increased from 3.02 to 3.38. The difference was statistically significant, *MD* = 0.68, *SE* = 0.29, *t*(129.80) = 2.34, *p* = .02103, 95% CI [0.10, 1.26], Hedges' *g* = 0.38. This improvement suggests that students became more aware of honesty, transparency, and personal responsibility when using AI.

Learner agency also improved more strongly in the experimental group. The mean score increased from 3.08 to 4.02, compared with an increase from 3.06 to 3.41 in the control group. The difference was significant, *MD* = 0.59, *SE* = 0.26, *t*(130.12) = 2.31, *p* = .02238, 95% CI [0.09, 1.10], Hedges' *g* = 0.38. This finding indicates that students became more capable of making decisions, revising their own work, evaluating AI-generated suggestions, and taking responsibility for the final text. In this sense, the intervention did more than improve scores; it helped students become more active authors of their own learning.

The detailed dimensions of academic integrity and learner agency further support this finding. In academic integrity, the highest mean score was for awareness of plagiarism, followed by transparency and honesty. In learner agency, the highest scores were observed for decision-making and responsibility. These dimensions indicate that students became more aware of ethical AI use and more capable of evaluating, accepting, rejecting, and revising AI-generated suggestions after the training.

Table 3 Academic Integrity and Learner Agency Dimensions in the Experimental Group

Variable	Dimension	Post-test Mean
Academic integrity	Honesty	4.10
Academic integrity	Transparency	4.15
Academic integrity	Responsibility	4.08
Academic integrity	Ethical AI use	4.05
Academic integrity	Awareness of plagiarism	4.20
Academic integrity	Ownership of writing	3.95
Learner agency	Goal setting	3.95
Learner agency	Self-regulation	4.00
Learner agency	Decision-making	4.10
Learner agency	Self-revision	4.05
Learner agency	Confidence	3.92
Learner agency	Responsibility	4.10

Students and lecturers also responded positively to the implementation of Contextual Writing Assessment. The mean perception score of students was 4.12, while that of lecturers

was 4.25. The highest score appeared in the relevance dimension, showing that both students and lecturers considered the assessment model suitable for English learning in the Faculty of Engineering. The lowest score was in the challenge dimension, indicating that the model required additional time, documentation, and classroom management.

Table 4 Mean Scores of Students' and Lecturers' Perceptions on Contextual Writing Assessment

Dimension	Student's Mean	Lecturers' Mean
Usefulness	4.20	4.30
Practicality	3.95	4.10
Relevance	4.25	4.40
Challenge	3.75	3.90
Acceptance	4.15	4.25
Sustainability	4.05	4.20
Overall mean	4.12	4.25

The implementation of Contextual Writing Assessment was also consistent. The overall treatment fidelity score was 86%, exceeding the minimum criterion of 80%. The highest implementation score appeared in the final contextual writing submission, while the lowest appeared in AI-use documentation. This indicates that students were generally able to complete the final contextual writing task, although documenting AI use required more guidance.

Table 5 Treatment Fidelity Score

Syntax of Contextual Writing Assessment	Implementation Percentage
Contextual problem orientation	88%
Personal idea generation	85%
Ethical AI support	87%
AI-use documentation	84%
Drafting and feedback	86%
Revision and reflection	85%
Final contextual writing submission	90%
Overall mean	86%

Qualitative Findings

The qualitative findings explain how Contextual Writing Assessment influenced students' writing behavior. Data from AI-use logs showed that students mainly used AI for brainstorming, vocabulary support, grammar checking, sentence improvement, and clarity. The overall AI-use log score was 3.30 out of 4.00, indicating good transparency in the documentation of AI use. Students were generally able to mention the AI tool used, explain the purpose of use, write prompts, summarize AI outputs, identify accepted and rejected suggestions, explain personal revisions, provide academic reasons, and write ethical statements.

The reflective journals also showed that the students became more aware of their writing processes. The overall reflective journal score was 3.28 of 4.00. Students were able to explain

their initial ideas, describe how AI was used, identify accepted and rejected AI suggestions, explain independent revisions, describe their learning experiences, and explain how originality was maintained. These findings indicate that students did not merely use AI to produce final texts; they reflected on how AI supported or influenced their writing decisions.

Draft analysis further confirmed the development of the students' writing process. A total of 90% of the students submitted initial drafts, 85% submitted revised drafts, and 100% submitted final writing products. In addition, 78% of the students made meaningful revisions, 76% showed personal contribution, and 72% revised beyond grammar correction. These results indicate that the intervention shifted students' attention from final product submission to process-based writing.

Table 6 Summary of Qualitative Evidence

Data Source	Main Indicator	Result
AI-use log	Overall mean score	3.30/4.00
Reflective journal	Overall mean score	3.28/4.00
Draft analysis	Students submitted initial drafts	90%
Draft analysis	Students submitted revised drafts	85%
Draft analysis	Students submitted final products	100%
Draft analysis	Students showed meaningful revision	78%
Draft analysis	Students showed personal contribution	76%
Draft analysis	Students revised beyond grammar correction	72%

The interview data supported the quantitative and document-based findings. Among the 18 selected student informants, 15 indicated awareness of responsible AI use, 14 reported reduced copy-paste behavior, 13 reported improved self-revision, 12 reported increased writing confidence, 14 reported better contextual awareness, 15 showed awareness of academic integrity, and 15 were able to explain their AI use. These findings indicate that most students became more selective and responsible when using AI for writing.

The lecturer interviews also showed positive responses. All four lecturers stated that the model was relevant to English for Engineering and improved the transparency of AI use. Three lecturers stated that the model helped reduce AI copy-paste behavior, was practical to implement, and was suitable for future use. Two lecturers noted that the model required technical simplification, especially in managing AI-use logs and reflective notes.

Table 7 Interview Themes

Informant Group	Theme	Frequency
Students	Responsible AI use	15/18
Students	Reduced copy-paste behavior	14/18
Students	Improved self-revision	13/18
Students	Increased writing confidence	12/18
Students	Better contextual awareness	14/18
Students	Awareness of academic integrity	15/18
Students	Ability to explain AI use	15/18
Lecturers	Model relevant to English for Engineering	4/4

Informant Group	Theme	Frequency
Lecturers	Model helped reduce AI copy-paste behavior	3/4
Lecturers	Model improved transparency in AI use	4/4
Lecturers	Model was practical to implement	3/4
Lecturers	Model required technical simplification	2/4
Lecturers	Model was suitable for future use	3/4

Overall, the findings show that the Contextual Writing Assessment was effective in reducing dependence on AI-generated text. The quantitative data indicated a stronger decrease in AI dependence and stronger improvement in writing performance, originality, academic integrity, and learner agency in the experimental group. Qualitative data explained that these improvements occurred because students were required to document AI use, revise their writing, reflect on their decisions, and demonstrate personal responsibility for the final product. Thus, the intervention shifted writing assessment from final-product evaluation to contextual, process-based, and ethically accountable assessments.

Discussion

The central finding is that Contextual Writing Assessment reduced students' dependence on AI-generated text more effectively than conventional writing assessment. The experimental group showed a mean reduction of 1.28 points, compared with 0.35 points in the control group. The between-group difference was statistically significant, $t(128.39) = -2.44$, $p = .01595$, Cohen's $d = -0.399$, with an approximate $R^2 = .044$. Thus, group membership accounted for about 4.4% of the variance in change scores. Although this magnitude is small to moderate, the direction and consistency of the result are educationally meaningful because the intervention addressed not only whether students used AI, but how they documented, evaluated, revised, and took responsibility for AI-assisted writing.

This result extends the argument that assessment redesign is more constructive than relying solely on prohibition or detection. (Gonsalves, 2025) noted that conventional product-based assessment is increasingly vulnerable because generative AI can produce coherent and apparently acceptable texts. The present study adds classroom-based evidence that contextual tasks, draft histories, AI-use logs, reflective journals, and ethical statements can restore process visibility and strengthen assessment validity. In this model, AI was not treated as misconduct by default; instead, students were required to demonstrate the reasoning and revision that transformed technological assistance into an accountable learning process.

The originality findings reinforce this interpretation. The experimental group improved significantly more than the control group in writing originality, $t(130.00) = 2.32$, $p = .02186$, $d = 0.379$, $R^2 = .040$. A complementary pattern appeared in similarity scores, which declined more strongly in the experimental group, $t(130.08) = -2.31$, $p = .02221$, $d = -0.378$, $R^2 = .040$. These converging indicators are important because originality and similarity are related but not identical constructs. Higher originality suggests stronger personal contribution and contextual reasoning, whereas lower similarity indicates reduced textual overlap. Together,

they suggest that students moved beyond surface-level paraphrasing and produced writing that more clearly reflected their own decisions and disciplinary context (Susilo et al., 2026).

Contextual Writing Assessment also generated a statistically reliable improvement in English writing performance. The difference in gain scores favored the experimental group, $t(129.96) = 2.32$, $p = .02170$, $d = 0.379$, $R^2 = .040$. The N-gain comparison showed the same pattern, $t(128.36) = 2.45$, $p = .01586$, $d = 0.399$, $R^2 = .045$, with the experimental group reaching a mean N-gain of 0.44 compared with 0.22 in the control group. These results indicate that reducing dependence on AI did not restrict students' access to useful support. Rather, the intervention appears to have redirected AI use toward brainstorming, vocabulary assistance, grammar checking, and clarity improvement while preserving the student's responsibility for content development and final textual decisions(Soodan et al., 2024; Sulistyanto et al., 2025).

The improvement in academic integrity was similarly significant, $t(129.80) = 2.34$, $p = .02103$, $d = 0.381$, $R^2 = .040$. This finding suggests that integrity can be strengthened through explicit instructional and assessment practices rather than through abstract warnings alone. Requiring students to disclose the tool used, explain the purpose of use, identify accepted and rejected suggestions, and justify revisions made ethical decision-making visible and assessable. The high post-test means for plagiarism awareness, transparency, honesty, and responsibility further indicate that students increasingly understood academic integrity as an active practice of disclosure, judgment, and ownership.

Learner agency also increased more strongly in the experimental group, $t(130.12) = 2.31$, $p = .02238$, $d = 0.377$, $R^2 = .039$. This result is theoretically important because dependence on AI-generated text is not only a technological problem; it is also a problem of self-regulation, confidence, and decision-making. Contextual Writing Assessment required students to generate initial ideas, evaluate AI suggestions, revise independently, and explain why particular changes were accepted. These practices positioned students as active authors who remained accountable for the final text, thereby supporting the decision-making, self-revision, and responsibility dimensions of learner agency.

Across the seven quantitative comparisons, Cohen's d ranged from |0.377| to |0.399|, while the approximate R^2 values ranged from .039 to .045. These statistics indicate small-to-moderate effects, not large intervention effects. This interpretation is important for scientific accuracy. The intervention produced a consistent pattern across multiple outcomes, but it did not explain most of the variance in students' development. Writing performance, integrity, agency, and AI dependence are shaped by prior proficiency, motivation, digital literacy, lecturer support, task difficulty, and other classroom factors. The value of the intervention therefore lies in the convergence and practical relevance of its effects rather than in an exaggerated claim of magnitude.

The qualitative evidence helps explain the mechanisms underlying these statistical differences. AI-use logs and reflective journals achieved mean scores of 3.30 and 3.28 out of 4.00, respectively, indicating relatively strong transparency and reflection. Moreover, 78% of students made meaningful revisions, 76% demonstrated personal contribution, and 72% revised beyond grammatical correction. Interviews showed that most student informants reported responsible AI use, reduced copy-paste behavior, stronger self-revision, and

greater awareness of academic integrity. This triangulation suggests that the quantitative improvements were accompanied by observable changes in writing behavior rather than being limited to score fluctuations.

Implementation evidence also supports the feasibility of the model. Treatment fidelity reached 86%, exceeding the predetermined criterion of 80%, and the final contextual writing submission achieved the highest implementation score. Students and lecturers reported positive overall perceptions, with mean scores of 4.12 and 4.25, respectively. However, these perception results should be interpreted descriptively, particularly because only four lecturers participated. Accordingly, H5 is best regarded as descriptively supported rather than as the result of a stable lecturer-level parametric test. The lower score for the challenge dimension also indicates that AI-use documentation and reflective requirements can increase workload and should be supported by concise templates and clear examples.

The findings contribute to writing assessment in STEM and engineering education. (Khuder & Negretti, 2025) emphasized that writing in STEM contexts requires regulation, metacognition, and instructional support. The present study shows that these principles can be operationalized through tasks that assess technical relevance, contextual appropriateness, revision history, ethical AI use, and personal reasoning. This contribution is particularly relevant to Indonesian higher education, where empirical studies connecting AI-resilient assessment, English for engineering, academic integrity, and learner agency remain limited.

Several limitations qualify the interpretation of the statistical findings. The study used six intact classes rather than individual random assignment, and the primary analyses were conducted at the student level. Because students were nested within classes, the reported standard errors may be somewhat optimistic, and the results should be confirmed using a larger number of clusters and multilevel analysis. The intervention was also implemented in one institution and one course, while some outcomes relied on self-report measures. In addition, the reported R^2 values are approximate variance-explained indices derived from the Welch statistics, not coefficients from a multivariable regression model. Future studies should report verified regression or multilevel-model R^2 values when examining predictors of reduced AI dependence.

Taken together, the findings support a human-centered interpretation of AI-assisted writing. Assessment remained focused on students' reasoning, disciplinary understanding, revision choices, and responsibility for the final text, while AI was retained as a limited support tool. Theoretically, the study strengthens the view that assessment validity in the generative-AI era depends on contextual relevance, process evidence, and ethical accountability. Practically, it offers lecturers a structured approach for reducing uncritical AI dependence without removing potentially useful forms of technological assistance.

Conclusion

This study demonstrates that Contextual Writing Assessment can reduce students' dependence on AI-generated text while supporting more responsible and independent writing. The experimental group achieved a significantly greater reduction in AI dependence than the control group, $t(128.39) = -2.44$, $p = .01595$, $d = -0.399$, $R^2 = .044$. This result supports H1 and indicates that requiring contextual reasoning, documented AI use, drafting, revision, and reflection can shift AI from a substitute for writing toward a controlled learning resource.

The intervention also produced significantly stronger gains in originality, $t(130.00) = 2.32$, $p = .02186$, $d = 0.379$, $R^2 = .040$, together with a greater reduction in similarity scores, $t(130.08) = -2.31$, $p = .02221$, $d = -0.378$, $R^2 = .040$. Writing performance improved significantly, $t(129.96) = 2.32$, $p = .02170$, $d = 0.379$, $R^2 = .040$, and the N-gain comparison confirmed the advantage of the intervention, $t(128.36) = 2.45$, $p = .01586$, $d = 0.399$, $R^2 = .045$. These findings support H2 and H3 and show that ethical control of AI use can coexist with measurable improvement in writing quality.

Contextual Writing Assessment further strengthened academic integrity, $t(129.80) = 2.34$, $p = .02103$, $d = 0.381$, $R^2 = .040$, and learner agency, $t(130.12) = 2.31$, $p = .02238$, $d = 0.377$, $R^2 = .039$. These results support H4 and indicate that students became more transparent in reporting AI use, more capable of evaluating AI-generated suggestions, and more responsible for revising and authorizing their final texts. Qualitative evidence from AI-use logs, reflective journals, draft analysis, and interviews corroborated these statistical patterns.

The effect sizes and variance-explained indices were consistently small to moderate, with |d| values of approximately 0.38-0.40 and approximate R^2 values of .039-.045. The intervention should therefore be interpreted as producing reliable and educationally meaningful advantages rather than dramatic or deterministic effects. H1-H4 were statistically supported, whereas H5 was descriptively supported by positive student and lecturer perceptions and an implementation-fidelity score of 86%. This balanced interpretation strengthens the credibility of the findings and avoids overstating the intervention's impact.

The study concludes that Contextual Writing Assessment offers a viable human-centered and AI-resilient approach to English writing assessment. Lecturers can apply the model by requiring contextualized tasks, initial and revised drafts, concise AI-use logs, reflective explanations, ethical statements, and evidence of personal contribution. Because the study involved intact classes in a single institutional context, future research should replicate the model across disciplines and institutions, use larger numbers of class clusters, employ multilevel or verified regression analyses, and examine the long-term development of academic integrity, learner agency, writing autonomy, and responsible AI use.

References

Abolina, A., Zorins, A., Vindaca, O., Usca, S., & Petrushenko, Y. (2026). Transforming Higher Education through AI-enriched Tools: an open innovation approach. *Frontiers in Education*, *11*(February). https://doi.org/10.3389/feduc.2026.1766004

Asrifan, A., Cardoso, L. M. O. de B., & Vargheese, K. J. (2025). Artificial intelligence in bilingual

ESP: A mixed-methods study on English for educational technology. *Journal of English Language Teaching Innovations and Materials*, *7*(1), 88–112. https://doi.org/10.26418/jeltim.v7i1.91274

Daniarti, Y., Firmansyah, M. A., & Alamsyah, D. (2026). Enhancing EFL Listening Through Structured Autonomous Use of British Council Learn English Audio Materials. *International Journal of Research in Education*, *6*(1), 96–106. https://doi.org/10.26877/ijre.v6i1.3274

Faridah, F., Luthfiyati, D., & Fatmawaty, R. (2026). Exploring Students ' acceptance of ChatGPT in Assisting EFL Writing Tasks : A technology acceptance model perspective. *Journal on English as a Foreign Language*, *16*(1), 356–379. https://doi.org/10.23971/jefl.v16i1.10378

Fauziah, F., Diana, N., Putri, S., & Fadhli, T. (2025). Student Engagement with Artificial Intelligence (AI) Chatbot in English Language Learning: Theory of Planned Behavior Perspective. *JEELS (Journal of English Education and Linguistics Studies)*, *12*(June 2025), 29. https://doi.org/10.30762/jeels.v12i2.4085

Febriyanti, Nurchurifiani, E., Hastomo, T., Wicaksono, A., & Nasution, S. S. (2026). AI Literacy of Pre-Service English Teachers: Proficiency, Strategies, and Barriers. *JEELS (Journal of English Education and Linguistics Studies) P-ISSN:*, *13*(April 2026), 32. https://doi.org/10.30762/jeels.v13i1.6293

Firdes, A. P. (2024). Students' Voices: Using ChatGPT for Learning English. *ELT-Lectura: Studies and Perspectives in English Language Teaching*, *12*(2), 125–138. https://doi.org//10.31849/9yvvmy21

Gonsalves, C. (2025). Contextual Assessment Design in the Age of Generative AI. *Journal of Learning Development in Higher Education*, *34*, 14. https://doi.org/10.47408/jldhe.vi34.1307

Gunawan, F. C., & Susanto, A. I. F. (2025). Investigating the Impacts of ChatGPT on Students Motivation in Learning English Grammar. *Journal of English Language Teaching and Learning (JETLE)*, *6*(2), 146–157. https://doi.org/10.18860/jetle.v6i2.32397

Hapsari, E. W., & Sudarman, S. (2025). AI-Powered Learning: Shaping the Future of Academic Writing in EFL Education in Indonesia. *JADEs Journal of Academia in English Education*, *6*(1), 114–127. https://doi.org/10.32505/jades.v6i1.10484

Hidayat, H., Yuliasr, I., Rukmini, D., & Widhiyanto. (2025). A Quasi-experimental Study of Inquiry Learning in Hybrid Environments: Impact on writing and self-regulated learning levels. *Journal of English Language Teaching Innovations and Materials*, *7*(October), 21–37. https://doi.org/10.26418/jeltim.v7i2.97081

Imran, M., Almusharraf, N., Sayed, M., Milana, A., & Abbasova, Y. (2024). Interactive Mobile Technologies Enhancing Learning Systems and Transforming Educational Paradigms. *International Journal of Interactive Mobile Technologies*, *18*(18), 135–154. https://doi.org/10.3991/ijim.v18i18.49143

Khuder, B., & Negretti, R. (2025). Collaborative Writing Regulation: a comparative case study of co-regulation and socially shared regulation in higher education. *Higher Education*, *91*(2), 677–699. https://doi.org/10.1007/s10734-025-01437-9

Lumintang, S. F., & Kuswandono, P. (2025). Indonesian EFL Lecturers' Positive Perceptions and Adaptive Strategies in Implementing Authentic Assessment for Sustainable Education: Mixed-methods. *Journal of English Language Teaching Innovation and Materials*, *7*(2), 38–59. https://doi.org/10.26418/jeltim.v7i2.96334

Pariyanto, & Tungka, N. F. (2025). ChatGPT as a Formative Feedback Tool: Improving Narrative Essay Writing among EFL Students. *EDULITICS JOURNAL*, *10*(1), 71–84. https://e-jurnal.unisda.ac.id/index.php/edulitic/article/view/8657

Rafida, T., Suwandi, S., & Ananda, R. (2024). EFL Students' Perception in Indonesia and Taiwan on Using Artificial Intelligence To Enhance Writing Skills. *Jurnal Ilmiah Peuradeun*, *12*(3), 987–1016. https://doi.org/10.26811/peuradeun.v12i3.1520

Rohadi, T., Imani, A., Sulaeman, D., Aminuddin, M., Geng, H., & Mohammad Ekramul Islam, K. (2025). Intentional and Developmental Learners' Agency in Deep Learning Processes during EFL Writing: An explanatory sequential design. *Journal of English Language Teaching Innovations and Materials (Jeltim)*, *7*(2), 1–20. https://doi.org/10.26418/jeltim.v7i2.97197

Septyaningrum, L., Wiyaka, W., & Prastikawati, E. F. (2024). Picture-Cued Sequence Task: How Does It Effect to Students' English Writing Ability? *International Journal of Research in Education*, *4*(2), 410–422. https://doi.org/10.26877/ijre.v4i2.816

Soodan, V., Rana, A., Jain, A., & Sharma, D. (2024). AI Chatbot Adoption in Academia: task fit, usefullness, and collegial ties. *Journal of Information Technology Education: Innovations in Practice*, *23*, 1–27. https://doi.org/10.28945/5260

Sulistyanto, I., Prasongko, A., & Tannaka, Z. M. (2025). Fostering EFL Students' Academic Literacy: Students' Perception using Elicit AI. *EDULITICS*, *10*(2), 2015. https://doi.org/10.52166/edulitics.v10i2.11361

Susilo, A., Firdaus, M., & Yukamana, H. (2026). AI Integration in English Language Teaching : Junior High School Teachers ' Perceptions and Readiness in Ogan Ilir , South Sumatera. *International Journal of Research in Education*, *6*(1), 62–71. https://doi.org/10.26877/ijre.v6i1.3108

Ulfah, B. (2026). From Convenience to Concern: Exploring AI Paraphasing Tools in Academic Writing Practise Across Indonesia. *EDULITICS JOURNAL*, *10*(2), 122–135. https://doi.org/10.52166/edulitics.v10i2.11300

Yunita, R., Fitriana, N., & Maulana, N. I. (2025). Artificial Intelligence in English Writing: Ethical Interation and Educators' Perspectives. *Journal of English Education and Linguistics Studies (JEELS)*, *12*(2), 26. https://doi.org/10.30762/jeels.v12i2.5250