Journal Homepage: -www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)

ISSN (O): 2320-5407
ISSN (P): 3107-4928

Article DOI: 10.21474/IJAR01/23392
DOI URL: http://dx.doi.org/10.21474/IJAR01/23392

RESEARCH ARTICLE

ASSESSING AI-SUPPORTED ACADEMIC WRITING: THRESHOLD PERFORMANCE, TRACEABILITY, AND PEDAGOGICAL MEDIATION IN A FIRST-YEAR UNIVERSITY COURSE

Héctor A. Aponte-Alequín

1. Division of Institutional Research and Assessment, University of Puerto Rico, Río Piedras Campus, 00925 San Juan, Puerto Rico.

Manuscript Info

Manuscript History
Received: April 2, 2026
Final Accepted: May 19, 2026
Published: May 30, 2026

Keywords:
Generative artificial intelligence; academic writing; learning assessment; pedagogical mediation; traceability; authorial voice; information literacy; critical thinking

Abstract

Background: Generative artificial intelligence (AI) has intensified debates about academic writing, authorship, and the validity of learning assessment evidence in higher education. This study examined how three pedagogical conditions of academic writing were associated with threshold performance, traceability, and pedagogical mediation in a first-year university Spanish course.

Methods: A mixed-methods design was used in a general education course at the University of Puerto Rico, Río Piedras Campus. The quantitative component analyzed 35 student writing performances distributed across three conditions: AI-integrated in situ writing, non-AI in situ writing, and non-AI take-home writing. Student performance was examined through an institutional analytic rubric aligned with Effective Communication, Information Literacy, and Critical Thinking. The qualitative component analyzed 12 AI writing logs to identify docu mented decisions of acceptance, rejection, reformulation, verification, and authorial control.

Results: The take home condition produced the highest overall perform ance. However, the AI-integrated in situ condition generated more balanced and traceable evidence of learning, meeting the institutional threshold in Effective Communication and Information Literacy. Neither in situ condition met the institutional threshold for Critical Thinking. Qualitative findings showed that AI writing logs documented lexical precision, textual cohesion, conceptual appropriation, and authorial judgment during revision.

Conclusion: AI-supported writing did not automatically produce stronger evidence of learning. Its educational value depended on structured pedagogical mediation, transparent documentation, and assessment conditions that made student decision-making visible. The study suggests that AI can support academic writing assessment when integrated through staged, traceable, and criteria-based instructional designs.

Corresponding Author: Héctor A. Aponte-Alequín
Address:-Division of Institutional Research and Assessment, University of Puerto Rico, Río Piedras Campus, 00925 San Juan, Puerto Rico.

Introduction:-

Generative artificial intelligence (AI) has rapidly altered the conditions under which academic writing is taught, produced, revised, and assessed in higher education. Its availability has intensified debates about authorship, academic integrity, student agency, and the validity of evidence used to document learning. Recent scholarship has moved beyond deficit-oriented views that treat AI only as a threat to academic honesty and has emphasized the need for pedagogical frameworks that clarify acceptable uses, instructional purposes, and assessment implications [1-4]. From this perspective, the central question is not whether AI should be permitted or prohibited in writing classrooms, but under what conditions its use can support learning while preserving transparency, authorship, and interpretive rigor.

This question is particularly relevant in first-year university writing courses, where students are expected to develop core academic literacies, including effective communication, information use, critical analysis, and control over formal written discourse. In these settings, AI tools may support revision, lexical precision, textual organization, and source-related practices. However, they may also obscure the extent to which the final text reflects student learning, external assistance, automated rewriting, or undocumented forms of support. As a result, AI-supported writing creates a methodological and pedagogical challenge for learning assessment: instructors and institutions must distinguish between improved textual products and evidence that can sustain valid inferences about what students know, can do, and can justify.

Learning assessment frameworks provide a useful lens for addressing this challenge. Assessment that supports learning requires explicit criteria, meaningful tasks, and evidence that documents performance in ways that are both interpretable and actionable [5,6]. Authentic assessment further emphasizes the importance of tasks that reflect meaningful academic practices and make student decision-making visible. Equity-oriented assessment also requires attention to the conditions under which evidence is produced, since unequal access to time, support, prior knowledge, or technological resources may shape performance in ways that are not always visible in the final product [7]. Rubrics and structured instructional scripts can help students regulate their work and make assessment criteria more transparent, but their value depends on how they are embedded in the learning process [8].In AI-mediated writing, traceability becomes especially important. Traceability refers here to the documentation of the student's writing and revision process, including the prompts used, AI-generated suggestions received, decisions accepted or rejected, revisions made, and reasons offered for those decisions. Such documentation can transform AI use from an opaque external support into an assessable component of the learning process. Writing logs, drafts, guided prompts, and reflective annotations can provide evidence of authorial control, verification, reformulation, and metacognitive awareness [9-11]. Without such mechanisms, a polished final text may reveal little about the student's actual learning, judgment, or participation in the writing process.

Research on AI writing tools also suggests that students may perceive these technologies as useful for improving clarity, formal register, organization, and linguistic accuracy [12]. Yet these potential benefits coexist with concerns about voice, stylistic homogenization, and overreliance on automated language [13]. In academic writing, authorial voice is not merely a matter of personal style. It also involves the ability to make purposeful lexical, rhetorical, conceptual, and evidentiary choices in relation to audience, genre, and disciplinary expectations. Therefore, responsible AI integration in writing instruction must help students evaluate suggestions, verify information, preserve intention, and justify revisions. AI can support writing development when it is integrated within a pedagogical design that requires students to remain accountable for the final text and the decisions behind it [14].

The institutional context of this study adds another layer to this problem. The University of Puerto Rico, Río Piedras Campus has adopted institutional guidelines for the integration and use of AI in academic and research projects and has also implemented a learning assessment framework organized around institutional learning domains and achievement thresholds [15,16]. Within this framework, the Online Learning Assessment System (OLAS) supports the documentation and analysis of student learning evidence through rubrics aligned with institutional domains. For writing-intensive tasks, this structure allows instructors to examine student performance in domains such as Effective Communication, Information Literacy, and Critical Thinking. However, the integration of AI raises an important question for institutional assessment: should writing produced under different conditions be interpreted as equivalent evidence of learning?This issue is especially complex in Puerto Rico, where university writing instruction takes place within a sociolinguistic ecology shaped by Spanish, English, academic language expectations, and broader questions of linguistic legitimacy [17]. In such a context, AI-supported writing cannot be reduced to a technical matter of grammar correction or textual improvement. It also involves questions of voice, linguistic security, register, and students' ability to maintain control over meaning while negotiating academic norms. For first-year students, these issues are central to their development as university writers.This article builds on a previous exploratory phase that examined AI writing logs as formative and assessment-oriented instruments in a university Spanish writing course [18,19]. That earlier phase suggested that structured AI use could support textual revision, metacognitive reflection, and authorial decision-making when students were required to compare drafts, document decisions, and rewrite in controlled conditions. The present study extends that line of inquiry by comparing three pedagogical conditions of academic writing in a first-year general education course: AI-integrated in situ writing, non-AI in situ writing, and non-AI take-home writing [20]. This comparison makes it possible to examine how different authentic writing conditions produced different kinds of assessment evidence, including final-product performance and, in the AI-integrated condition, documented evidence of student decision-making generated through the required writing log.

The study addresses a gap in the literature and in institutional practice. Although many discussions of AI in education focus on policy, ethics, instructional innovation, or student perceptions, fewer studies examine how AI-supported writing interacts with threshold-based learning assessment across institutional domains. In particular, there is a need to understand how performance varies across writing conditions and how student documentation of AI use can help interpret that performance. This distinction is essential because high scores alone do not necessarily provide the strongest evidence of learning if the production conditions are less controlled or less transparent.Accordingly, this study examined how three pedagogical conditions of academic writing were associated with threshold performance, traceability, and pedagogical mediation in a first-year university course. The quantitative component analyzed student performance through an institutional analytic rubric aligned with Effective Communication, Information Literacy, and Critical Thinking. The qualitative component examined AI writing logs to identify documented decisions of acceptance, rejection, reformulation, verification, and authorial control. Together, these forms of evidence allowed the study to consider both observed achievement and the quality of the evidence supporting its interpretation.

The study was guided by the following research questions:

1. How did student performance vary across the three pedagogical writing conditions in the domains of Effective Communication, Information Literacy, and Critical Thinking?
2. Which conditions met the institutional achievement threshold in each domain?
3. What types of revision, verification, reformulation, and authorial decisions did students document in AI writing logs?
4. How can quantitative performance results and qualitative evidence of traceability be integrated to interpret the pedagogical value of AI-supported academic writing?

By addressing these questions, the article contributes to ongoing discussions about AI integration, academic writing instruction, and learning assessment in higher education. Its central argument is that the educational value of AI-supported writing does not reside in the tool itself, nor in the final text alone. Rather, it depends on the pedagogical architecture that makes student decision-making visible, aligns writing tasks with explicit criteria, and supports careful interpretation of evidence across assessment conditions.

Conceptual Framework:

This study is grounded in a conceptual framework that connects learning assessment, AI-supported academic writing, and pedagogical mediation. The framework assumes that student writing should not be interpreted only as a final textual product, but also as evidence of decisions, revisions, verification practices, and authorial control. This assumption is especially relevant in AI-mediated contexts, where the apparent quality of a final text may not fully reveal how the student used available support or what learning can be inferred from the task.The first concept guiding the study is threshold performance. In institutional learning assessment, thresholds provide a reference point for determining whether student performance meets an expected level of achievement in a given domain. In this study, threshold performance is not treated as a purely numerical outcome. Rather, it is interpreted as an institutional signal that must be read in relation to the conditions under which the evidence was produced. This distinction is important because a high-performing text generated under less controlled conditions may not provide the same type of evidence as a text produced through a structured and documented process. Assessment that matters requires attention not only to scores, but also to the interpretive quality of the evidence supporting those scores [5-7].The second concept is traceability.

Traceability refers to the documentation of the writing process in ways that make student decision-making visible. In AI-supported academic writing, traceability may include the initial draft, the prompt used, the AI-generated response, the student's comparison of versions, accepted and rejected suggestions, reformulated passages, verification practices, and reflective explanations. This documentation is central because AI can improve surface features of writing while also making the source of those improvements less visible. Writing logs, therefore, can function as both pedagogical and assessment instruments: they help students articulate what they did, why they did it, and how the revision relates to academic criteria [9-11].

The third concept is pedagogical mediation. AI tools do not operate pedagogically by themselves. Their educational value depends on the instructional architecture that frames their use. Pedagogical mediation includes the teacher's decisions about task design, sequencing, prompts, permitted uses of AI, documentation requirements, rubric alignment, and opportunities for rewriting. In this study, mediation is understood as the process through which AI use is transformed from a possible shortcut into a structured learning experience. This view is consistent with research that emphasizes the role of explicit criteria, guided revision, and instructional scaffolding in helping students regulate their writing and make informed decisions [4,8].These three concepts are closely related. Threshold performance identifies whether students reached an expected level of achievement. Traceability helps explain how that performance was produced. Pedagogical mediation clarifies the instructional conditions that made the evidence more or less interpretable. Together, they support a more nuanced interpretation of AI-supported writing than a simple comparison between AI and non-AI conditions. The framework therefore shifts attention from the tool itself to the conditions that organize its use.

The framework also incorporates authorial voice as a key dimension of academic writing. In AI-supported contexts, voice refers not only to stylistic individuality, but also to the student's capacity to preserve purpose, select appropriate wording, evaluate suggestions, and maintain control over meaning. AI-generated or AI-revised language may enhance clarity and formal register, but it may also introduce homogenized phrasing or alter the writer's intended emphasis. For this reason, the study treats authorial voice as an observable dimension of student decision-making rather than as an abstract personal trait [12-14]. When students explain why they accept, reject, or modify AI suggestions, they provide evidence of voice as academic agency.Finally, the framework is situated within an institutional assessment context. The study examines student writing in relation to three learning domains: Effective Communication, Information Literacy, and Critical Thinking. These domains allow the analysis to move beyond general impressions of writing quality and toward a more specific interpretation of what types of learning were demonstrated. Effective Communication is associated with clarity, organization, linguistic accuracy, and formal academic expression. Information Literacy is associated with citation practices, verification, and responsible use of information. Critical Thinking is associated with interpretation, conceptual integration, and analytical depth. In AI-supported writing, these domains may not develop uniformly. A tool may support clarity and citation practices more readily than complex literary interpretation or critical analysis.

Accordingly, this study uses the conceptual framework of threshold performance, traceability, pedagogical mediation, and authorial voice to interpret evidence across three writing conditions. The purpose is not to determine whether AI is inherently beneficial or harmful, but to examine how different task conditions produce different types of evidence. This framework prepares the methodological design that follows: a mixed-methods analysis of rubric-based performance across institutional domains, complemented by qualitative analysis of AI writing logs as evidence of student decision-making.

Methods:-

Research Design:

This study used an exploratory mixed-methods comparative design to examine how three pedagogical conditions of academic writing were associated with student performance, traceability, and pedagogical mediation in a first-year university course. The quantitative component focused on rubric-based student performance across three institutional learning domains: Effective Communication, Information Literacy, and Critical Thinking. The qualitative component examined AI writing logs as process-oriented evidence of student decision-making during revision[24].A mixed-methods design was appropriate because the study did not seek only to compare scores across conditions. It also aimed to interpret how students documented revision, verification, reformulation, and authorial control when generative AI was integrated into the writing process. Therefore, quantitative results were used to identify patterns of threshold performance, while qualitative evidence was used to explain how traceability and mediation shaped the interpretive value of those results [21-23].The design was implemented in a real course context. For that reason, the study should not be understood as a randomized experiment. The three writing conditions differed in instructional support, time available for writing, and degree of process documentation. These differences were central to the study because the purpose was to examine how task conditions shaped not only performance, but also the quality of evidence available for learning assessment.

Setting:

The study was conducted at the University of Puerto Rico, Río Piedras Campus, in ESPA 3003, Fundamentals of Language and Discourse. This course is part of the undergraduate general education component and supports the development of academic writing, textual organization, discourse analysis, and formal written communication in Spanish. Because of its curricular location, the course provides a relevant setting for examining how first-year students produce academic texts and how their work can be assessed through institutional learning domains.The institutional context was also relevant to the study design. The campus uses the Online Learning Assessment System (OLAS) to document learning assessment results through rubrics aligned with institutional domains. In this study, OLAS provided the basis for organizing and interpreting student performance in Effective Communication, Information Literacy, and Critical Thinking. The study was also framed by institutional guidelines for responsible AI integration in academic and research projects [15,16].

Participants and Sample:

The sample consisted of 35 student writing performances produced by students enrolled in ESPA 3003. These performances were distributed across three pedagogical writing conditions. Eleven performances corresponded to the AI-integrated in situ condition, 13 to the non-AI in situ condition, and 11 to the non-AI take-home condition.

The qualitative component included 12 AI writing logs produced by students in the AI-integrated condition. These logs were selected because they documented how students interacted with AI-generated suggestions, what revisions they accepted or rejected, and how they justified their decisions. The qualitative sample was not intended for statistical generalization. Rather, it served as process-oriented evidence for interpreting how students made visible their revision practices, verification strategies, and authorial control.

Pedagogical Writing Conditions:

The intervention compared three pedagogical writing conditions. These conditions differed in the degree of AI integration, instructional mediation, time available, and process traceability.

Condition 1: AI-integrated in situ writing:

Students in this condition completed a staged writing sequence that integrated Microsoft Copilot as a guided support tool. The sequence included initial writing without AI, interaction with Copilot through a teacher-provided prompt, comparison between the student draft and the AI-supported version, completion of an AI writing log, and final in-class rewriting based on the student's own notes. AI use was not authorized as a substitute for authorship. Instead, it was framed as a bounded support for revision, verification, and reflection.The instructor-provided prompt used with Microsoft Copilot was the following: "A continuación, tecomparto un borrador de unareflexión que tengoasignada para un curso de españoluniversitario. Por favor, mejórala sin indicarmecuálesfueronloscambios que integraste. (Below, I'm sharing a draft of a reflection I have been assigned for a university Spanish course. Please improve it without telling me what changes you made.)" This prompt was intentionally constrained because the purpose of the task was not to outsource writing decisions to the tool, but to require students to compare the AI-supported version with their own draft and then document acceptance, rejection, or reformulation decisions in the writing log.

Condition 2: Non-AI in situ writing:

Students in this condition completed the writing task in class without the integrated AI model. This condition offered a more controlled writing environment, but without the structured AI-supported revision process or the additional time available in the take-home condition.

Condition 3: Non-AI take-home writing:

Students in this condition completed the writing task outside the classroom, without the integrated AI model. This condition allowed more time for elaboration, but provided less immediate control over the writing environment and fewer mechanisms for documenting the writing process.These three conditions allowed the study to compare not only AI and non-AI writing, but also different configurations of mediation, time, control, and traceability.Table 1 summarizes the three pedagogical writing conditions and the main evidence sources used in the study.

Table 1
Pedagogical writing conditions and evidence sources

Condition	N	Writing context	AI integration	Main evidence source
AI-integrated in situ writing	11	In-class staged writing and rewriting	Guided use of Microsoft Copilot through teacher-provided prompt and AI writing log	Final text, rubric scores, AI writing logs
Non-AI in situ writing	13	In-class writing	No integrated AI model	Final text and rubric scores
Non-AI take-home writing	11	Out-of-class writing	No integrated AI model	Final text and rubric scores

Note: AI = artificial intelligence

Instruments and Data Sources:
Four main sources of evidence were used: written reflections, an institutional analytic rubric, OLAS records, and AI writing logs.The written reflection was the academic writing task assessed in the study. Students were asked to produce a brief academic text in Spanish that included literary interpretation, formal written expression, conceptual integration, and the use of references. The task required students to demonstrate writing competencies that could be aligned with Effective Communication, Information Literacy, and Critical Thinking.The institutional analytic rubric was used to assess student performance. The rubric used an eight-point scale organized into four performance levels: Beginning (1-2), Developing (3-4), Good (5-6), and Excellent (7-8). The institutional threshold was defined as at least 70% of students achieving a score of 5 or higher in the criteria associated with a given learning domain. The rubric included criteria related to literary analysis, clarity, morphosyntactic accuracy, citations and references, essay structure, integration of theories and conceptual frameworks, development of social issues, and growth in analytical complexity. For this article, the analysis focused on the criteria aligned with Effective Communication, Information Literacy, and Critical Thinking.

OLAS records provided the quantitative data for the study. These records made it possible to organize performance by rubric criterion, domain, and pedagogical condition. Because OLAS is the institutional platform used to document student learning assessment, it allowed the course-level evidence to be interpreted within a broader institutional assessment framework.The AI writing logs provided qualitative evidence of traceability. In these logs, students documented original fragments, AI-supported revisions, aspects of the text that improved, explanations based on course concepts, rejected or modified suggestions, and lessons learned from the revision process. The logs were used to examine how students made visible their decisions of acceptance, rejection, reformulation, verification, and authorial control.The full institutional rubric and the AI writing log template are available upon request from the corresponding author at hector.aponte2@upr.edu.

Procedure:
Data collection occurred as part of the regular instructional sequence of the course. Students completed writing tasks under one of the three pedagogical conditions described above. In the AI-integrated condition, the process followed a staged sequence: initial writing without AI, guided interaction with Microsoft Copilot, comparison of versions, completion of the AI writing log, and in-class rewriting. This sequence was designed to preserve student authorship while making the revision process visible.In the non-AI in situ condition, students completed the writing task in class without the integrated AI sequence. In the non-AI take-home condition, students completed the writing task outside class. In both non-AI conditions, the final written product served as the main evidence of performance.After the tasks were completed, student writing was assessed using the institutional analytic rubric, and the results were organized in OLAS. AI writing logs from the AI-integrated condition were then analyzed qualitatively to identify patterns of revision, verification, reformulation, and authorial decision-making.

Quantitative Analysis:
The quantitative analysis was descriptive and comparative. Student performance was examined by pedagogical condition and institutional learning domain. For each condition, the analysis considered mean performance and the percentage of students who met the institutional achievement threshold.The threshold was defined as 70% of students achieving a score of 5 or higher on the rubric criteria associated with each domain. Results were interpreted by domain rather than only as global writing scores. This decision was important because AI-supported revision may affect different dimensions of writing unevenly. For example, it may support clarity, structure, or citation practices

more readily than deep critical analysis.Given the exploratory nature of the study and the small sample size, the quantitative analysis did not aim to establish causal effects or statistical generalization. Instead, it was used to identify performance patterns across authentic instructional conditions and to support a cautious interpretation of learning assessment evidence.However, to support scoring consistency, inter-rater reliability was examined through the intraclass correlation coefficient. The rubric scoring yielded an ICC of 0.81, which is considered excellent according to the interpretive guidelines proposed by Koo and Li (2016). This result supported the use of the rubric scores as sufficiently consistent for the exploratory descriptive analysis conducted in this study.

Qualitative Analysis:
The qualitative analysis focused on the 12 AI writing logs from the AI-integrated condition. The logs were analyzed through a combination of thematic analysis and qualitative content analysis [22,23]. The purpose was to identify recurrent patterns in how students documented their use of AI and justified their writing decisions.The analysis considered segments in which students described what they accepted, rejected, modified, verified, or learned during the revision process. Coding was primarily inductive, although it was informed by the conceptual framework of traceability, pedagogical mediation, and authorial voice. Special attention was given to evidence of lexical precision, cohesion, formal academic register, conceptual appropriation, verification practices, and student control over meaning.The qualitative phase was not treated as an illustrative supplement to the quantitative results. Rather, it provided metaevidence for interpreting how performance was produced under the AI-integrated condition. This evidence was essential for examining whether the writing process itself became visible and assessable.

Ethical Considerations:
The study was conducted as an institutional learning assessment project within a real course context. The writing tasks analyzed in the study were used for formative and institutional interpretation, not to assign summative consequences across unequal writing conditions. This decision was important because the three groups worked under different conditions of time, support, and process documentation.AI use was explicitly delimited in the AI-integrated condition. Students were required to document their interaction with the tool, preserve evidence of revisions, and justify decisions. This requirement supported transparency and reduced the risk of treating AI-generated language as unexamined student authorship.Student data were analyzed through institutional records and anonymized writing logs. The analysis was conducted with attention to confidentiality, responsible interpretation, and the limits of evidence produced in a non-randomized course setting. Therefore, findings are presented as contextual and transferable by analogy, rather than as generalizable causal claims.

Results:

Overall Performance by Pedagogical Writing Condition:
The quantitative results showed clear differences across the three pedagogical writing conditions. The non-AI take-home condition produced the highest overall performance, with a mean score of 86.93%. The AI-integrated in situ condition followed, with a mean score of 74.43%. The non-AI in situ condition produced the lowest overall performance, with a mean score of 50.48%. The same pattern was observed when the percentage of students reaching the institutional threshold was considered. In the non-AI take-home condition, 100% of students reached the expected level of performance. In the AI-integrated in situ condition, 81.82% reached the threshold. In contrast, only 7.69% of students in the non-AI in situ condition reached the expected level. These results suggest a consistent performance order: non-AI take-home writing, followed by AI-integrated in situ writing, followed by non-AI in situ writing.Table 2 presents the overall performance by pedagogical writing condition.

Table 2
Overall performance by pedagogical writing condition

Pedagogical writing condition	N	Overall mean (%)	Students ≥70% (%)	Threshold met
AI-integrated in situ writing	11	74.43	81.82	Yes
Non-AI in situ writing	13	50.48	7.69	No
Non-AI take-home writing	11	86.93	100.00	Yes

Note: AI = artificial intelligence

Although the take-home condition produced the highest quantitative result, the three conditions were not equivalent in terms of time, control, mediation, or traceability. Therefore, these results are presented descriptively and should not be interpreted as evidence of causal effects.

Performance by Institutional Learning Domain:

When student performance was examined by institutional learning domain, the overall pattern remained consistent. The non-AI take-home condition produced the highest scores in Effective Communication, Information Literacy, and Critical Thinking. The AI-integrated in situ condition showed intermediate performance and met the institutional threshold in two of the three domains. The non-AI in situ condition showed the lowest performance and did not meet the threshold in any domain.In Effective Communication, the AI-integrated in situ condition obtained a mean score of 6.33 out of 8, equivalent to 79%. The non-AI in situ condition obtained a mean score of 3.89, equivalent to 49%. The non-AI take-home condition obtained a mean score of 7.33, equivalent to 92%. Both the AI-integrated in situ and the non-AI take-home conditions met the institutional threshold in this domain.

In Information Literacy, the AI-integrated in situ condition obtained a mean score of 6.27, equivalent to 78%. The non-AI in situ condition obtained a mean score of 3.92, equivalent to 49%. The non-AI take-home condition obtained a mean score of 7.27, equivalent to 91%. Again, both the AI-integrated in situ condition and the non-AI take-home condition met the institutional threshold.In Critical Thinking, the differences between conditions were less pronounced, but the pattern remained the same. The AI-integrated in situ condition obtained a mean score of 5.18, equivalent to 65%. The non-AI in situ condition obtained a mean score of 4.54, equivalent to 57%. The non-AI take-home condition obtained a mean score of 6.18, equivalent to 77%. Only the non-AI take-home condition met the institutional threshold in Critical Thinking.Table 3 presents domain-level performance and threshold status across conditions.

Table 3
Domain-level performance and threshold status by pedagogical writing condition

Institutional learning domain	AI-integrated in situ writing Mean / % / Threshold	Non-AI in situ writing Mean / % / Threshold	Non-AI take-home writing Mean / % / Threshold
Effective Communication	6.33 / 79% / Yes	3.89 / 49% / No	7.33 / 92% / Yes
Information Literacy	6.27 / 78% / Yes	3.92 / 49% / No	7.27 / 91% / Yes
Critical Thinking	5.18 / 65% / No	4.54 / 57% / No	6.18 / 77% / Yes

Note: Values are presented as mean score out of 8 / percentage / threshold status.
AI = artificial intelligence.

Because the study was exploratory and the conditions were not experimentally controlled, these domain-level patterns should be interpreted as descriptive and hypothesis-generating. They suggest possible relationships among time, task setting, AI mediation, documentation, and domain-level performance, but they do not establish causal effects.These results show that the AI-integrated in situ condition performed above the institutional threshold in Effective Communication and Information Literacy, but not in Critical Thinking. This finding is important because it suggests that AI-supported writing may relate differently to specific dimensions of learning. In this study, the guided AI condition was associated with stronger evidence in domains related to textual clarity, organization, citation practices, and information use than in the domain requiring deeper analytical interpretation.

Qualitative Findings from AI Writing Logs:

The qualitative analysis of 12 AI writing logs identified four recurring categories: lexical precision and academic register, cohesion and textual reorganization, appropriation of conceptual frameworks and disciplinary metalanguage, and authorial judgment in response to AI suggestions. These categories show that students did not merely document whether AI-generated suggestions improved their texts. They also explained how they evaluated, modified, rejected, or learned from those suggestions.Table 4 summarizes the qualitative categories identified in the AI writing logs.

Table 4
Qualitative categories from AI writing logs

Category	**Description**	**Representative translated excerpt**
Lexical precision and academic register	Students reported learning to avoid imprecise wording and select more formal or conceptually accurate language.	"I learned that in academic writing it is important to avoid imprecise words and use verbs that provide greater formality and conceptual accuracy."
Cohesion, clarity, and textual reorganization	Students identified how connectors, sentence order, and paragraph structure affected readability and meaning.	"I noticed that small changes in connectors and expressions can transform a long and confusing paragraph into a fluent and understandable text."
Appropriation of conceptual frameworks and disciplinary metalanguage	Students used course concepts to explain why a revision improved the text.	"The use of a formal register strengthens the credibility of the discourse."
Authorial judgment in response to AI suggestions	Students rejected or modified AI suggestions when they altered meaning, genre, emphasis, or authorial intention.	"I rejected changing 'poem' to 'text' because the original fragment refers to a specific literary genre."

The first category, lexical precision and academic register, appeared when students described how AI-supported revision helped them recognize vague wording and consider more formal alternatives. These comments suggest that students used AI output as a point of comparison for refining diction, register, and conceptual accuracy. In this sense, AI-supported revision contributed to visible reflection about academic language.The second category, cohesion, clarity, and textual reorganization, showed that students became more aware of relationships between sentence order, connectors, paragraph flow, and reader comprehension. The logs indicated that students did not interpret revision only as correction of errors. They also associated revision with the organization of meaning and the communicative effectiveness of the text.

The third category, appropriation of conceptual frameworks and disciplinary metalanguage, reflected students' use of course concepts to justify revisions. Students referred to communication principles, textuality, formal register, clarity, and semiotic interpretation when explaining changes. This category is important because it shows a connection between revision practices and disciplinary learning. The writing log required students not only to state what changed, but also to explain why the change mattered in relation to concepts studied in the course.The fourth category, authorial judgment in response to AI suggestions, provided evidence that students did not accept all AI-generated revisions automatically. Some students rejected lexical substitutions because they weakened genre-specific meaning, altered emphasis, or did not fit the intended context. Others modified AI suggestions to preserve their own interpretation or improve concision. These entries suggest that the AI writing log made authorial control visible as a form of decision-making.

Integration of Quantitative and Qualitative Results:

The integration of quantitative and qualitative evidence provides a more nuanced view of the AI-integrated in situ condition. Quantitatively, this condition did not produce the highest overall performance. The non-AI take-home condition obtained higher scores in all three domains. However, the AI-integrated in situ condition produced a combination of satisfactory performance and process documentation that was not available in the other two conditions.The qualitative evidence helps interpret this pattern. The AI writing logs showed that students documented specific revision decisions, used academic and disciplinary concepts to justify changes, and sometimes rejected AI suggestions. Therefore, the AI-integrated condition generated evidence not only of the final written product, but also of the process through which students revised and controlled that product.

At the same time, the results show an important limitation. The AI-integrated in situ condition met the institutional threshold in Effective Communication and Information Literacy, but not in Critical Thinking. The qualitative logs contained evidence of authorial judgment and metacognitive reflection, but this evidence did not translate into threshold-level performance for the group in Critical Thinking. This suggests that AI-supported revision and documentation may support textual improvement and information-related practices more readily than deeper analytical development.Overall, the results indicate that different writing conditions produced different types of assessment evidence. The non-AI take-home condition produced the highest quantitative performance. The AI-integrated in situ condition produced more traceable evidence of student decision-making. The non-AI in situ condition produced the weakest performance across domains. These differences support the need to interpret writing assessment results in relation to task conditions, rather than treating all written products as equivalent evidence of learning.

Discussion:-

The findings of this study show that AI-supported academic writing cannot be interpreted through a simple comparison between AI and non-AI conditions. The results reveal a more complex pattern: the non-AI take-home condition produced the highest quantitative performance, whereas the AI-integrated in situ condition generated the most pedagogically traceable and institutionally interpretable evidence of learning. This distinction is central to the study's contribution. In learning assessment, the strongest score is not always the strongest evidence. A high-performing text may demonstrate a polished product, but it does not necessarily reveal how the student produced it, what decisions shaped it, or what forms of support influenced the final version.The non-AI take-home condition clearly produced the highest overall performance. Students in this group reached the highest mean score and met the institutional threshold in Effective Communication, Information Literacy, and Critical Thinking. This outcome should not be minimized. It suggests that additional time, access to resources, and opportunities for revision outside the classroom can support stronger academic writing. In practical terms, the take-home condition allowed students to produce more complete, organized, and analytically developed texts than those produced in the non-AI in situ condition. However, the same features that may have supported higher performance also complicate the interpretation of the evidence. Because the writing process occurred outside the classroom and without systematic documentation, the final product offers limited information about the sources of improvement, the extent of student independence, or the specific decisions involved in revision.

This issue is not merely procedural. It is central to assessment validity. Authentic assessment requires evidence that can support meaningful inferences about what students know, can do, and can justify [5,6]. When a writing task is completed under less controlled conditions, the final text may reflect student learning, but it may also reflect undocumented assistance, unequal access to support, or differences in time and resources. From an equity-oriented assessment perspective, these conditions matter because they affect the interpretive quality of the evidence [7]. Therefore, although the take-home condition produced the strongest quantitative outcome, it did not necessarily produce the most transparent or pedagogically interpretable evidence of learning.The AI-integrated in situ condition offers a different type of contribution. It did not outperform the take-home condition, but it met the institutional threshold in Effective Communication and Information Literacy and produced documented evidence of revision decisions. This contribution should not be attributed to AI alone. Rather, it emerged from the instructional sequence that combined initial writing, a constrained teacher-provided prompt, comparison of versions, an AI writing log, rubric alignment, and controlled rewriting. Under these conditions, the writing log functioned as a component of the AI-integrated assessment design, not as an independent outcome.

It allowed students to document lexical decisions, cohesion adjustments, verification practices, and authorial judgment in ways that final texts alone would not reveal.The Critical Thinking results require particular caution. The AI-integrated in situ condition did not meet the institutional threshold in this domain, whereas the take-home condition did. One cautious interpretation is that the mediated AI sequence may have supported clarity, organization, formal register, and information-related practices more readily than deeper analytical development. However, this interpretation remains speculative because the observed pattern may also reflect differences in time available, task setting, sample size, prior preparation, or the constrained nature of the teacher-provided prompt. Future iterations of the model should therefore incorporate stronger supports for critical interpretation, including structured prewriting, guided close reading, explicit modeling of analytical reasoning, and opportunities to revise conceptual depth rather than textual form alone.The contrast between the two in situ conditions suggests that control over the writing environment is not sufficient by itself. In this study, the condition that combined control with structured mediation and documentation produced more useful assessment evidence than the condition based only on in-class writing.The contrast between the AI-integrated condition and the non-AI in situ condition further clarifies the role of mediation. Both conditions occurred in a more controlled environment than the take-home condition. However, the non-AI in situ group performed substantially lower across domains. This pattern suggests that control alone does not produce strong evidence of learning. A controlled task without adequate scaffolding may reduce the possibility of undocumented support, but it may also restrict students' ability to demonstrate their best writing. The AI-integrated condition, by contrast, combined control with structured support. Students worked under mediated conditions that provided revision assistance, documentation requirements, and opportunities for reflection.

Thus, the difference between these two in situ groups highlights that pedagogical mediation matters as much as the presence or absence of AI.

As a cautious institutional implication, assessment systems that include AI-mediated writing may benefit from recording basic information about task conditions. Such documentation could indicate whether AI use was permitted, guided, prohibited, or required; whether the task was completed in class or outside class; and whether students submitted drafts, logs, or reflective annotations. The present study does not establish the scalability or feasibility of this practice across programs or disciplines. It suggests, however, that documenting task conditions may help instructors and assessment offices interpret threshold results more responsibly.Accordingly, institutions that assess writing in AI-mediated contexts should consider documenting task conditions as part of the assessment record. This documentation may include whether AI use was permitted, prohibited, guided, or required; whether the task was completed in class or outside class; whether students submitted drafts or writing logs; and whether revision decisions were documented. Such information would allow instructors, programs, and assessment offices to interpret threshold performance more responsibly. It would also help distinguish between performance as product and learning as demonstrated process.

Taken together, the findings support a cautious distinction between performance as product and assessment evidence as process. The take-home condition produced the strongest observed performance, while the AI-integrated condition provided additional documentation of how students engaged with revision. For learning assessment, both forms of evidence are valuable, but they support different kinds of interpretation.Overall, the discussion supports the central argument of the study: the educational value of AI-supported academic writing depends on the quality of the pedagogical and assessment design. The take-home group "wins" in observed performance because it produced the highest scores. The AI-integrated condition should therefore not be interpreted as evidence that AI alone produced greater traceability. Rather, it illustrates how an AI-supported task, when designed with required documentation, guided prompting, rubric alignment, and controlled rewriting, can generate a different kind of assessment evidence than final-product writing alone.For learning assessment, both findings matter. Strong writing performance is important, but so is the ability to explain how that performance was achieved. In AI-mediated academic writing, the most useful evidence may not be the most polished text, but the text whose production process can be examined, justified, and connected to explicit learning criteria.Consequently, the findings invite a shift in how educators and institutions approach AI in writing instruction. The question should not be whether AI improves writing in general. The more precise question is: under what instructional conditions can AI-supported writing generate evidence of learning that is strong, fair, traceable, and useful for improvement? This study suggests that the answer lies in staged, criteria-based, and documented pedagogical designs. When AI is integrated through mediation, writing logs, rubric alignment, and opportunities for controlled rewriting, it can contribute not only to better texts, but also to better evidence of learning.

Limitations:-

This study has several limitations. First, the sample was small and drawn from one first-year course at a single institution. Second, the design was exploratory and non-randomized, so the findings should not be interpreted as causal effects of AI use. Third, the three writing conditions differed in time, support, control, and documentation, which was analytically relevant but limits direct comparability. Fourth, the qualitative analysis was based on 12 AI writing logs, whose depth depended on students' ability and willingness to document their decisions. Finally, the assessment relied on an institutional rubric applied in a formative context, which may have shaped student motivation and performance.All interpretations regarding Critical Thinking should therefore be read as speculative and hypothesis-generating rather than as evidence of a differential effect of AI on this domain.

Conclusion:-

This study examined AI-supported academic writing through the lens of learning assessment, threshold performance, traceability, and pedagogical mediation. The findings show that different writing conditions produced different types of evidence. The non-AI take-home condition yielded the highest quantitative performance and met the institutional threshold in Effective Communication, Information Literacy, and Critical Thinking; however, it offered limited evidence about how the final texts were produced. By contrast, the AI-integrated in situ condition did not reach the highest overall scores, but it met the institutional threshold in Effective Communication and Information Literacy and generated documented evidence of student decision-making through AI writing logs. These logs made visible processes often hidden in final written products, including lexical selection, textual reorganization, verification, rejection of AI suggestions, and authorial control. Thus, the study suggests that AI-supported writing should not be evaluated only by whether it produces higher scores or more polished texts, but by whether it creates fair, traceable, and pedagogically meaningful evidence of learning.

At the same time, the AI-integrated in situ condition did not meet the institutional threshold in Critical Thinking, which indicates that AI-supported revision may strengthen clarity, formal register, organization, and information-related practices more readily than deeper analytical reasoning. AI integration should therefore not be understood as a substitute for intensive teacher mediation, conceptual modeling, close reading, discussion, and iterative analytical practice. Future research should examine this model with larger samples, additional course sections, and more controlled comparisons across writing conditions. Further studies should also explore how AI writing logs can be refined to better capture critical thinking, conceptual integration, and authorial voice.

The long-term effects of implementing this AI-integrated writing model should be examined beyond immediate task performance. If sustained across multiple writing assignments, course sections, and academic terms, the model may help students develop more stable habits of revision, verification, and authorial decision-making. It may also support a gradual shift from product-centered assessment toward process-sensitive evidence of learning, especially when writing logs, guided prompts, rubric alignment, and controlled rewriting are used consistently. However, these potential effects remain exploratory and should be tested through longitudinal designs that examine whether students transfer these practices to new writing contexts, non-AI tasks, and more advanced disciplinary assignments. Future research should therefore investigate not only whether AI-supported writing improves textual quality, but also whether mediated and traceable AI use strengthens students' long-term writing autonomy, critical judgment, and confidence as academic writers.Overall, the central conclusion of this study is that AI-supported academic writing is neither inherently beneficial nor inherently detrimental to learning assessment. Its value depends on transparent, criteria-based, and traceable pedagogical design.

Acknowledgments:

The author acknowledges the institutional context of the Division of Institutional Research and Assessment at the University of Puerto Rico, Rio Piedras Campus, where the assessment work that informed this study was developed.

Data Availability Statement:

The institutional rubric and AI writing log template used in this study are available from the corresponding author upon reasonable request. Student-level data are not publicly available due to confidentiality considerations.

References:-

1. Bearman M, Ryan J, Ajjawi R. Discourses of artificial intelligence in highereducation: A criticalliteraturereview. HigherEducationResearch&Development 2023. doi:10.1080/07294360.2023.2193178.
2. Chan CKY. A comprehensive AI policyeducationframeworkforuniversityteaching and learning. International Journal of EducationalTechnology in HigherEducation2023;20:38. doi:10.1186/s41239-023-00391-4.
3. Noroozi O, Soleimani S, Farrokhnia M, et al. Generative AI in education: Pedagogical, theoretical, and methodologicalperspectives. International Journal of Technology in Education 2024;7(3):373–385. doi:10.46328/ijte.845.
4. Perkins M, Furze L, Roe J, et al. The Artificial IntelligenceAssessmentScale (AIAS): A frameworkforethicalintegration of generative AI in educationalassessment. Journal of UniversityTeaching and LearningPractice 2024;21(6). doi:10.53761/q3azde36.
5. Jankowski NA. Movingtoward a philosophy of assessment. AssessmentUpdate 2017;29(3):10–11. doi:10.1002/au.30096.
6. Jankowski NA, Timmer JD, Kinzie J, et al. AssessmentThatMatters: TrendingTowardPracticesThatDocumentAuthenticStudentLearning; NationalInstituteforLearningOutcomesAssessment: Urbana, IL, USA, 2018.
7. Montenegro E, Jankowski NA. A New DecadeforAssessment: EmbeddingEquityintoAssessment Praxis; NationalInstituteforLearningOutcomesAssessment: Urbana, IL, USA, 2020.
8. Fraile J, Gil-Izquierdo M, Medina E. Theimpact of rubrics and scripts onself-regulation, self-efficacy, and performance in collaborativeproblem-solvingtasks. Assessment&Evaluation in HigherEducation 2023;48(8):1223–1239. doi:10.1080/02602938.2023.2236335.

9. Shibani A, Buckingham Shum S. AI-assistedwriting in education: Ecosystemrisks and mitigations. Proceedings of the In2Writing '24: TheThird Workshop onIntelligent and InteractiveWritingAssistants, Honolulu, HI, USA,
10. May 2024. Associationfor Computing Machinery: New York, NY, USA, 2024. doi:10.1145/3690712.3690714.
11. UNESCO. Guía para el uso de IA generativa en educación e investigación. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000389227 (accessedon 12 May 2026).
12. U.S. Department of Education, Office of EducationalTechnology. EmpoweringEducationLeaders: A ToolkitforSafe, Ethical, and Equitable AI Integration. Available at: https://files.eric.ed.gov/fulltext/ED661924.pdf (accessedon 12 May 2026).
13. Gasaymeh A-MM, Beirat MA, Abu Qbeita AA. Universitystudents' insights of generative artificial intelligence (AI) writingtools. EducationSciences 2024;14(10):1062. doi:10.3390/educsci14101062.
14. Sandstead M, Kibler AK. Voice in L2 writing in theage of AI. Journal of SecondLanguageWriting2025;69:101212. doi:10.1016/j.jslw.2025.101212.
15. Zamorano C. Enhancingacademicwriting in English languageeducationthroughgenerative AI integration. ResearchStudies in English LanguageTeaching and Learning 2025;3(3):424–447. doi:10.62583/rseltl.v3i3.87.
16. Decanato de Asuntos Académicos. Circular 8 (2024–2025): Lineamientos y guías para la integración y uso de la inteligencia artificial (IA) en los proyectos académicos y de investigación del Recinto de Río Piedras; Universidad de Puerto Rico, Recinto de Río Piedras: San Juan, Puerto Rico, 2025.
17. Maldonado Vargas E, Montañez García A, Aponte-Alequín HA, et al. Plan Institucional de Avalúo del Aprendizaje Estudiantil 2025–2029; División de Investigación Institucional y Avalúo, Universidad de Puerto Rico, Recinto de Río Piedras: San Juan, Puerto Rico, 2025. Available at: https://academicos.uprrp.edu/diia/wp-content/uploads/sites/5/2025/09/PIAAE-24sept2025.pdf (accessedon 12 May 2026).
18. Morales Lugo K, Carroll KS, Campos SD. Introduction to specialissue: Languages in modernday Puerto Rico. International Journal of theSociology of Language 2024;2024(286):1–22. doi:10.1515/ijsl-2023-0061.
19. Aponte-Alequín HA. Enhancingthewritingprocess: Integratingappliedlinguisticslearningassessment in theclassroom. EuropeanScientificJournal, ESJ 2025;21(5):35. doi:10.19044/esj.2025.v21n5p35.
20. Aponte-Alequín HA. Inteligencia artificial en la enseñanza de redacción: Oportunidades de comunicación efectiva y pensamiento crítico; División de Investigación Institucional y Avalúo, Universidad de Puerto Rico, Recinto de Río Piedras: San Juan, Puerto Rico, 2025. Available at: https://academicos.uprrp.edu/diia/wp-content/uploads/sites/5/2025/11/InvestIAExploratorio.pdf (accessedon 12 May 2026).
21. Aponte-Alequín HA. Inteligencia artificial y avalúo del aprendizaje: Análisis del umbral en los dominios de Competencias de Información, Comunicación Efectiva y Pensamiento Crítico en una muestra exploratoria; División de Investigación Institucional y Avalúo, Universidad de Puerto Rico, Recinto de Río Piedras: San Juan, Puerto Rico, 2026. Unpublishedinstitutionalresearchreport.
22. Creswell JW, Creswell JD. ResearchDesign: Qualitative, Quantitative, and MixedMethodsApproaches, 5th ed.; SAGE: ThousandOaks, CA, USA, 2018.
23. Braun V, Clarke V. ThematicAnalysis: A PracticalGuide; SAGE Publications Ltd.: London, UnitedKingdom, 2021.
24. Mayring P. Qualitative Content Analysis: A Step-by-StepGuide; SAGE Publications Ltd.: London, UnitedKingdom, 2022. doi:10.4135/9781036231798.
25. Koo TK, Li MY. A guideline of selecting and reportingintraclasscorrelationcoefficientsforreliabilityresearch. Journal of Chiropractic Medicine 2016;15(2):155–163. doi:10.1016/j.jcm.2016.02.012.