MSTS Methods in Science and Technology Studies
futuristech PT. Teknologi Futuristik Indonesia
Date of publication April 23, 2026, date of current version April 23, 2026.
Digital Object Identifier 10.64539/msts.v2i1.2026.415

Article

Impact of Generative AI on Student Learning in Higher Education using Robust Assessment Metrics Framework

Adesola M. Falade[1], Ayobami E. Mesioye[2,*]

[1] Department of Software Engineering, McPherson University, Seriki Sotayo, Ogun State, Nigeria; e-mail: faladeam@mcu.edu.ng (A. M. Falade).
[2] Department of Cybersecurity, McPherson University, Seriki Sotayo, Ogun State, Nigeria; e-mail: mesioyeae@mcu.edu.ng (A. E. Mesioye).
* Correspondence Author

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Abstract: The rapid emergence of Generative Artificial Intelligence (GAI) has transformed the landscape of higher education, influencing pedagogy, assessment, and student learning experiences. Despite its widespread adoption, a significant research gap persists regarding the empirical measurement of its impact on specific learning outcomes. While GAI tools are widely adopted, existing assessment frameworks often fail to distinguish between machine-generated efficiency and genuine cognitive development. This study addresses this gap by developing the Robust Assessment Metrics Framework (RAMF), evaluated through a mixed-methods approach involving students and faculty (N=295) at McPherson University. Quantitative findings reveal a significant "Efficiency-Cognition Trade-off": while frequent GAI usage significantly enhances task efficiency ($p < 0.001$), it correlates with a statistically significant decline in critical thinking ($p < 0.01$) and self-reported originality ($p < 0.01$). Interestingly, regression analysis shows that AI literacy and institutional policy clarity are stronger predictors of academic confidence than usage frequency. This suggests a psychological "confidence paradox" where students feel more capable despite lower cognitive engagement. Qualitatively, thematic analysis highlights a shift toward "shortcut learning" that necessitates a move from product-oriented to process-oriented evaluation. The RAMF introduces expert-validated protocols such as the ‘30/70 Synthesis Rule’ and "Process Logs," to safeguard academic rigor. This research provides institutional leaders with an expert-validated framework proposed for institutional trial to shift from product-oriented to process-oriented assessment in the AI era. By focusing on the interplay between human agency and algorithmic assistance, this research offers broader implications for pedagogical redesign in an AI-saturated academic environment.

Keywords: Learning outcomes; Assessment metrics; Pedagogy; Generative Artificial Intelligence (GAI), Cognitive engagement; Academic integrity.

Copyright: © 2026 by the authors. This is an open-access article under the CC-BY-SA license.

1. Introduction

The rapid spread of Generative AI (GAI) platforms, such as ChatGPT, Claude, and Gemini has profoundly impacted the educational landscape. Unlike earlier AI models that focused on pattern recognition, GAI generates human-like text, images, and software code. This offers a dynamic and interactive resource for both learners and educators [1], [2]. In higher education, these tools have become integral to academic practices, ranging from essay drafting and coding assistance to idea generation and personalized tutoring.

However, this integration represents a pivotal shift in teaching and learning methodologies. While GAI enhances creativity and efficiency, it simultaneously raises critical concerns regarding academic integrity, equity, and the quality of student learning [1], [2]. Educational institutions and policymakers are increasingly challenged to evaluate the effectiveness of these tools while addressing ethical implications. Recent studies emphasize the necessity of assessing student outcomes related to GAI usage [3], while others highlight the importance of understanding students’ familiarity with these tools [4]. Furthermore, the risk of bias within AI-based tools presents additional hurdles for marginalized student populations, necessitating a framework that aligns AI capabilities with ethical educational practices [5].

This increasing reliance on GAI intersects with global debates on teaching quality. While advocates argue that GAI democratizes access to knowledge, critics warn of overdependence and the potential erosion of critical thinking. At the center of this debate is the "Learning Outcome," yet traditional evaluation methods, such as standardized written assignments, have proven inadequate in capturing the full impact of GAI. A critical dilemma remains unresolved: traditional assessment focuses on the final "product," which GAI can now replicate with high proficiency, thereby masking the actual cognitive engagement of the student. This renders traditional metrics obsolete and creates an empirical gap in how we measure internal mastery.

This study addresses this empirical gap by proposing and evaluating the Robust Assessment Metrics Framework (RAMF), an expert-validated framework proposed for institutional trial. Specifically, it investigates how usage frequency influences the trade-off between efficiency and cognition, and how institutional factors predict student confidence.

The remainder of this paper is organized as follows: Section 2 reviews the existing literature and theoretical foundations. Section 3 details the mixed-methods methodology, including the institutional context and the development of the RAMF. Section 4 presents the quantitative and qualitative results, integrating them to explain the observed pedagogical paradoxes. Finally, Section 5 concludes the study with practical implications for academic leadership and suggestions for future research.

2. Literature Review

Generative Artificial Intelligence (GAI) has emerged as a disruptive force in higher education. It reshapes student learning, faculty curriculum design, and institutional evaluation of academic performance. Unlike earlier forms of educational technology, GAI mimics human creativity through the generation of original text, visual outputs, and interactive responses.

2.1. Generative AI in Higher Education

AI use in education has moved from adaptive learning platforms to advanced large language models. Research suggests that GAI can serve as a "cognitive partner" that provides real-time feedback and scaffolding [6]. However, scholars warn of "shortcut learning," a situation where students bypass intellectual engagement [7]. This tension underscores the need for assessments that differentiate between machine-assisted efficiency and genuine cognitive growth. Table 1 provides a narrative synthesis of the current state of research.

2.2. Theoretical Foundations and the Tool-Replacement Spectrum

The integration of GAI into pedagogy can be mapped along a spectrum ranging from augmentative tools to substitutive replacements. From a Social Constructivist perspective, learning occurs within the Zone of Proximal Development (ZPD). When GAI acts as a tool, it provides the scaffolding necessary for students to reach higher levels of synthesis [6]. However, the risk of "replacement" arises when GAI performs the cognitive labor entirely. If the student uses GAI to bypass the "productive struggle" required to construct meaning, the construction of knowledge is aborted. The RAMF framework targets this by measuring the process of inquiry rather than just the output. Furthermore, Self-Determination Theory (SDT) posits that for deep learning to occur, students need autonomy and competence [8]. When GAI is used for brainstorming, it can enhance a student's sense of competence. Conversely, generating an entire paper erodes academic self-efficacy and long-term mastery [9] This study aligns these theoretical lenses with the three domains of learning: the cognitive domain (targeting higher-order thinking), the affective domain (measuring engagement and ethical agency), and the psychomotor domain (evaluating technical fluency as digital craftsmanship) [10].

The relationship between these theoretical domains and the proposed assessment structure is visualized in Figure 1. This conceptual model illustrates how the RAMF operates at the intersection of cognitive processing, ethical agency, and technical productivity, ensuring that GAI remains a scaffold for student growth rather than a substitute for intellectual labor.

2.3. Research Gap

While the literature acknowledges the transformative ability of GAI, there exists a significant gap in the implementation of multidimensional assessment models. This research fills this void by developing an expert-validated conceptual model that balances efficiency gains with the protection of critical thinking, ensuring that GAI supports constructivist learning rather than replacing it.

3. Methodology

3.1. Research Design

This study utilizes a convergent mixed-methods design [11]. Data were collected from McPherson University, a private institution in Nigeria, targeting a diverse sample of students and faculty. This single-institution focus allows for a deep understanding of contextual factors, such as specific institutional policies, though it is acknowledged as a limitation regarding broader generalizability.

To ensure a rigorous and transparent methodology, the study followed a five-stage research flow as depicted

Table 1. Summary of Related Research on GAI in Higher Education.

Author	Research Title	Key Findings	Gaps / Limitations
[12]	Assessment and Learning Outcomes for GAI: A Scoping Review	Identifies three educator approaches: traditional, refocused, and GenAI-incorporated.	Purely descriptive; does not validate practical, multidimensional metrics.
[13]	Systematic Review of GAI for Teaching and Learning	Reviews GAI's role across various educational contexts.	Primarily conceptual synthesis; no assessment framework proposed.
[14]	ChatGPT's Impact on Data Science Learning	Provides support in coding but risks student dependence.	Single-discipline focus; lacks generalizable assessment metrics.
[15]	Impact of ChatGPT on Academic Integrity	ChatGPT disrupts integrity; signals need for assessment redesign.	Theoretical overview only; no empirical data or tested metrics.
[16]	Student and Teacher Perspectives on ChatGPT	Explores stakeholder views on benefits and concerns.	Perspective-focused; no concrete metrics for impact measurement.
[17]	Does ChatGPT Enhance Learning? Meta-Analysis	Improves performance and motivation but notes methodological weaknesses.	Studies lack power analysis and long-term originality measures.
[18]	Student & Staff Perceptions of AI in Assessment	Mixed feelings on GAI feedback; low familiarity among staff.	Perception-focused; lacks outcome-based metrics.
[19]	Could ChatGPT Get an Engineering Degree?	GAI can pass standard exams (65-85% accuracy).	Measures GAI capability, not student learning outcomes.
[20]	AI Reshaping the Classroom (News Reports)	AI boosts short-term performance but may impair critical thinking.	Journalistic commentary; lacks empirical study data.

Table 2. Participant Demographics (N = 295).

Category	Sub-category	Frequency (*n*)	Percentage (%)
Participant Role	Students	248	84.1
	Faculty Members	47	15.9
Academic Discipline	STEM	103	34.9
	Social Sciences	104	35.3
	Humanities	88	29.8
Gender	Female	159	53.9
	Male	135	45.8
	Other/No Response	1	0.3

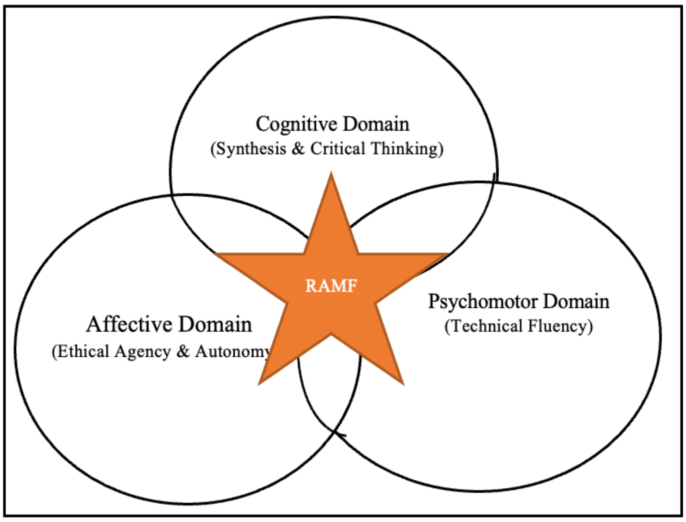

Figure 1. RAMF Conceptual Assessment Model.

in Figure 2. This process began with the conceptualization of the RAMF and proceeded through expert validation (CVI) before empirical data collection and eventual triangulation of findings.

3.2. Instrument Development and Validation

The survey instrument focused on three core constructs to assess perceived student engagement with GAI. To measure Critical Thinking, a five-item scale was utilized, featuring prompts such as *"I use GAI to compare multiple perspectives on a topic rather than accepting a single answer"* and assessments of whether students actively look for logical fallacies or "hallucinations" in GAI-generated text.

The dimension of Originality was operationalized through a four-item scale, including statements such as *"The final version of my AI-assisted assignments reflects my unique personal voice"* and *"I use GAI only for structural drafting, ensuring the core arguments are my own."* Finally, Academic Confidence was evaluated using four items, such as "I feel confident that I can achieve learning objec-

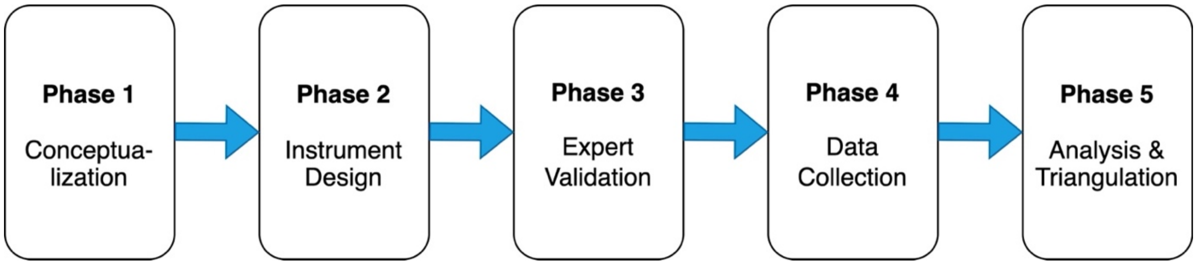

Figure 2. Research Flow Stage Flowchart.

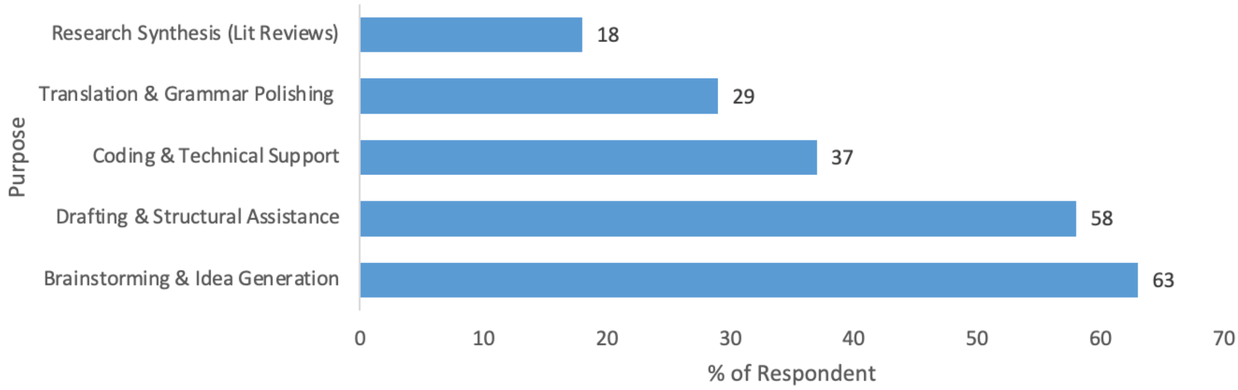

Figure 3. Primary Purposes for GAI Usage among Students (%).

Table 3. GAI Usage Patterns and Academic Purpose (N = 248 students).

Primary Use Case	Frequency (%)	Mean Satisfaction (1.00 – 5.00)
Brainstorming & Idea Generation	63	4.50
Drafting & Structural Assistance	58	4.20
Coding & Technical Support	37	4.60
Translation & Grammar Polishing	29	4.10
Research Synthesis (Lit Reviews)	18	3.20

tives while using GAI ethically," to capture the student's sense of self-efficacy in an AI-integrated environment.

Qualitative protocols were similarly structured; semi-structured interview guides were peer-reviewed by two educational technologists to ensure questions were non-leading and aligned with the study's theoretical framework of Constructivism and Self-Determination Theory [21].

3.3. Participants and Sampling

A purposive sampling strategy was employed across three academic clusters. Table 2 details the demographics. To ensure data integrity, the gender counts were audited; the sample includes 159 female, 135 male, and 1 student who identified as "Other/Prefer not to say," totaling 294 participants who disclosed gender out of the 295 total sample size.

3.4. Data Collection Procedures

Data were collected over 12 weeks during the 2024/2025 academic year. Surveys were administered electronically via institutional mailing lists. Focus groups (n =22 students) and semi-structured interviews (n = 5 faculty) were conducted via video conferencing, recorded, and transcribed verbatim.

3.5. Data Analysis Plan and Qualitative Rigor

Quantitative data were analyzed using SPSS (v.28). The tool utilizes independent samples t-tests to compare frequent versus low GAI users and multiple linear regression to identify predictors of academic confidence. For the qualitative component, thematic analysis followed the established six-step framework by [22]. Also, a multi-stage process was employed to ensure qualitative rigor and inter-rater reliability (IRR).

At the initial phase of independent coding, two researchers separately analyzed 20% of the transcripts, which included four interviews and one focus group. This was followed by a comparison and reconciliation phase, where the coders met to harmonize the initial codebook and discuss discrepancies specifically related to the "Critical Thinking" and "Originality" themes until a consensus was reached. The final step involved the calculation of inter-rater reliability, which yielded a Cohen's

Table 4. Comparative Analysis of Perceived Learning Outcomes.

Metric	Frequent Users (Mean)	Low Users (Mean)	t-value	p-value
Task Efficiency	4.32	3.51	7.12	< 0.001
Critical Thinking	3.10	3.65	-4.85	< 0.01
Originality	2.94	3.72	-5.10	< 0.01

Table 5. Multiple Linear Regression - Predictors of Academic Confidence.

Predictor Variable	Beta (β)	t	Sig. (p)
Responsible AI Literacy	0.42	5.21	0.001
Frequency of GAI Use	-0.12	-1.14	0.254
Institutional Policy Clarity	0.35	4.88	0.001

Table 6. Content Validity Index (CVI) for RAMF Dimensions (n = 8).

Dimension	Relevance	Clarity	Feasibility	Item-CVI
Critical Thinking	4.0	3.8	3.5	0.94
Originality	3.8	3.5	3.2	0.88
Ethical Use	4.0	4.0	3.8	0.98
Scale-Level CVI	-	-	-	0.91

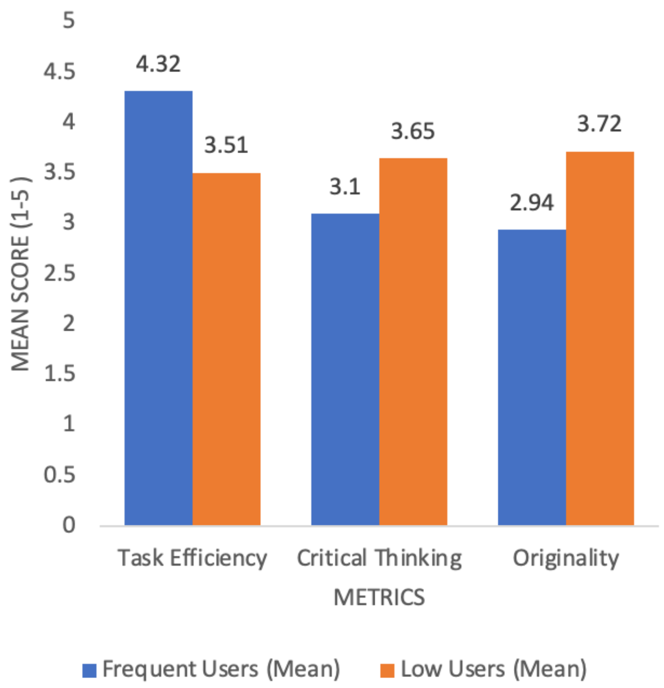

Figure 4. Impact of GAI Usage on Learning Outcomes.

Kappa of 0.82. This score is considered excellent and confirms the reliability of the thematic findings presented in the results section.

3.6. Integration and Triangulation

Following the convergent design, quantitative results (t-test) and qualitative themes ("AI-induced anxiety") were merged into a joint display [23]. This triangulation allows the researchers to explain *why* certain statistical trends, (such as the efficiency-depth trade-off) occur in the classroom environment.

3.7. Ethical Considerations

The study received approval from the McPherson University Research Ethics Committee. All participants were provided with a Plain Language Statement (PLS) and signed an electronic informed consent form. Data were stored on an encrypted server, and all identifying information was removed during the transcription of qualitative interviews.

4. Results and Discussion

4.1. Usage Patterns and Efficiency

The survey data collected from 248 students and 47 faculty members reveal that GAI is predominantly utilized as a structural scaffold rather than a primary research engine. As illustrated in Table 3 and visualized with explicit data labels in Figure 3, the highest usage frequency was recorded for brainstorming and idea generation at 63% (Mean Satisfaction = 4.50), followed by drafting and structural assistance at 58% (Mean Satisfaction = 4.20). Technical applications, such as coding and software support, showed the highest satisfaction level (Mean = 4.60), despite a lower usage frequency of 37%. Notably, research synthesis for literature reviews showed both the lowest usage (18%) and the lowest satisfaction (Mean = 3.20). This specific data point suggests that while students value the tool for organizational tasks, they remain skeptical of its scholarly accuracy, likely due to the documented risk of algorithmic hallucinations and inaccurate citations.

4.2. The Efficiency–Cognition Trade-off

To evaluate the impact of GAI usage frequency, an independent samples t-test was conducted to compare frequent users with low or non-users across key performance metrics of perceived learning. The results, summarized in Table 4 highlight a significant statistical divergence between productivity and engagement. As shown in the data labels for Figure 4, frequent users reported a high task efficiency mean of 4.32, which is significantly higher than the 3.51 reported by low users ($p < 0.001$). However, this gain in efficiency is countered by a decline in cognitive metrics.

Frequent users reported lower mean scores for perceived critical thinking (3.10) compared to low users (3.65, $p < 0.01$). A similar trend was observed for self-reported originality, where frequent users averaged 2.94 compared to 3.72 for their peers ($p < 0.01$). This empirical evidence confirms the "Efficiency-Cognition Trade-off," suggesting that while GAI accelerates the production of academic work, it may concurrently decelerate the deep processing of complex ideas.

4.3. Predictors of Confidence and "Confidence Paradox"

A multiple linear regression was performed to identify the primary predictors of student academic confidence. As shown in Table 5, the regression model explains 28% of the variance ($R^2 = 0.28$, $p < 0.001$). Interestingly, while usage frequency dictates the efficiency-cognition trade-off shown in Section 4.2, it fails to predict academic confidence in this regression model ($p = 0.254$). Instead, Responsible AI Literacy ($\beta = 0.42$, $p = 0.001$) and Institutional Policy Clarity ($\beta = 0.35$, $p = 0.001$) emerged as the only significant predictors.

This reveals a "Confidence Paradox" where students do not derive academic security from the mere frequency of tool usage. Rather, they feel most confident when they understand the ethical boundaries and technical logic of the tools. Without institutional transparency, frequent usage may actually lead to "misconduct anxiety," where high productivity is undermined by a fear of unintentional plagiarism. This confirms that confidence in the AI era is a psychological construct rooted in informed agency and institutional transparency, rather than technological reliance.

4.4. The Robust Assessment Metrics Framework (RAMF)

The RAMF is proposed as a multidimensional model to evaluate learning through the synergy between human cognition and AI assistance. It is important to clarify that while the framework's metrics and protocols have undergone content validation via expert consensus, the RAMF is currently an expert-validated framework proposed for institutional trial rather than an empirically tested utility. As shown in Table 6, the Scale-CVI was recorded at 0.91, which exceeds the threshold of 0.80 required for scientific validity. While experts noted that measuring "Originality" remains a subjective challenge (Feasibility = 3.20), the exceptionally high score for "Ethical Use" (0.98) confirms the framework's strength in promoting transparency and traceability.

4.5. Practical Application and Implementation Strategy

To provide faculty with actionable tools, the RAMF includes specific rubric descriptors designed to evaluate the "Originality and Voice" criterion. Under this scoring logic, an Exemplary performance (80–100%) is characterized by a clear "Human-Centric Narrative" where AI is limited to structural drafting, and 90% of the content includes personal reflection or localized data. A Proficient score (60–79%) is awarded when the student adheres to the 30/70 Synthesis Rule, ensuring that AI-generated definitions are cited correctly and edited to match an authentic academic voice. Conversely, work is classified as Emerging (40–59%) when there is an over-reliance on AI exceeding 50% of the text, resulting in a homogenized voice. Finally, a Fail grade (0–39%) is assigned to unedited GAI outputs that lack a "Prompt Log" or fail to demonstrate the "cognitive struggle" required for local contextualization.

The implementation of the RAMF is designed to be discipline-agnostic, allowing for customization across various academic clusters. In STEM fields, for instance, the required "Logic Log" may focus on the iterative debugging of AI-generated code, whereas in the Humanities, it centers on the refinement of philosophical arguments through iterative prompting. This specificity addresses the need for a concrete roadmap that moves from suspicion-based grading to evidence-based assessment. By documenting the "Prompt-to-Product" pipeline, institutions can finally measure the value added by the student versus the output generated by the machine, effectively resolving the pedagogical paradox identified in this study.

5. Conclusion and Limitations

5.1. Conclusion

The integration of Generative AI (GAI) into higher education is no longer a peripheral development but a central shift in the academic landscape. This study has demonstrated that GAI represents a "Pedagogical Paradox." While it offers unprecedented gains in task efficiency ($p < 0.001$), it poses a measurable risk to the development of critical thinking and original academic voice. The quantitative findings confirm that students who rely heavily on GAI without structured guidance report a significant decline in deep cognitive engagement.

However, the solution lies not in the prohibition of these tools, but in the evolution of assessment frameworks. The Robust Assessment Metrics Framework (RAMF) proposed in this study provides an expert-validated framework proposed for institutional trial for this transition. By shifting the focus from the 'Final Product' to the 'Cognitive Process' through the use of

Prompt Logs and the 30/70 Synthesis Rule, institutions can ensure that GAI serves as a 'thinking partner' rather than a 'proxy writer.

5.2. Practical Implications for Academic Leadership

For academic leadership, such as Deans and Department Heads, the implementation of the RAMF necessitates a strategic shift in institutional architecture. First, institutional leadership should prioritize policy standardization rather than relying on fragmented departmental rules. This can be achieved by establishing a university-wide "Traffic Light System"—categorizing tasks as Green (Full AI), Yellow (Assisted), or Red (No AI)—which aligns specifically with the RAMF's 30/70 synthesis rule. Additionally, faculty development must become a core focus of institutional investment. Dedicated "AI-Pedagogy Workshops" should be organized to train educators in the design of "AI-Resistant" prompts, specifically focusing on tasks that require localized, real-time, or deeply personal synthesis that GAI platforms currently cannot replicate.

Furthermore, resource allocation must reflect the shifting demands of process-based assessment. Academic leaders must recognize that evaluating iterative logic logs and conducting mini-vivas is significantly more labor-intensive than traditional product-oriented grading. To sustain these quality standards without overwhelming the workforce, it may be necessary to provide budgetary support for additional Teaching Assistants or to implement reduced credit-hour loads for faculty participating in RAMF-based curricula. By providing these structural and financial supports, institutions can transition from a posture of suspicion-based grading to a more robust, evidence-based model of quality assurance that treats GAI as a transparent tool for intellectual growth.

5.3. The "Implementation Burden": A Realistic Appraisal

While the RAMF ensures high academic integrity, it introduces a significant "Time and Complexity Premium." Reviewing iterative "Logic Logs" and conducting "Mini-Vivas" increases faculty workload by an estimated 15–20% compared to traditional essay grading. To mitigate this, institutions should adopt a Risk-Based Assessment Model, where the full RAMF protocols are applied to high-stakes summative work (for example, Final Dissertations), while lower-stakes formative tasks use simplified versions of the framework.

5.4. Limitations of the Study

Despite its contributions, this study is subject to several limitations. While the sample (N=295) provides a robust foundation, the data was collected from a limited number of institutions, which may restrict the generalizability of findings across different cultural or geographical contexts. Furthermore, the reliance on self-reported measures for critical thinking and originality introduces potential subjectivity and respondent bias. The rapid "technological velocity" of GAI evolution also means that observed impacts may shift as tools become more sophisticated in reasoning. Finally, the cross-sectional design offers only a "snapshot" of GAI's influence. Future longitudinal research is required to evaluate long-term effects on degree-level mastery.

5.5. Recommendations for Future Research

A critical priority for future investigation is the empirical field validation of the RAMF. While the current study establishes high content validity through expert review, controlled longitudinal trials are necessary to measure the framework's actual impact on objective student achievement and its feasibility in high-enrolment courses. Furthermore, scholars should investigate potential algorithmic biases in AI-assisted grading against English as a Second Language (ESL) or non-traditional writing styles to ensure assessment equity. Finally, comparative analyses between traditional detection-based "policing" models and integration-focused approaches, such as the RAMF, are essential to determine which strategy most effectively fosters long-term student motivation and integrity.

5. Declarations

5.1. Author Contributions

A. M. Falade: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Supervision**; A. E. Mesioye:** Formal analysis, Investigation, Resources, Data Curation, Writing - Review & Editing, Visualization, Project administration.

5.2. Institutional Review Board Statement

Approval was given.

5.3. Informed Consent Statement

An electronic consent form was signed by the participants.

5.4. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

5.5. Acknowledgment

Not applicable.

5.6. Conflicts of Interest

The authors declare no conflicts of interest.

6. References

[1] A. Yusuf, N. Pervin, and M. Román-González, "Generative AI and the future of higher education: a threat to academic integrity or reformation? Evidence from multicultural perspectives," *Int. J. Educ. Technol. High. Educ.*, vol. 21, no. 1, art. no. 21, 2024. https://doi.org/10.1186/s41239-024-00453-6.

[2] D. R. E. Cotton, P. A. Cotton, and J. R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," *Innovations in Education and Teaching International*, vol. 61, no. 2, pp. 228-239, 2023. https://doi.org/10.1080/14703297.2023.2190148.

[3] A. S. Nelson, P. V. Santamaría, J. S. Javens, and M. Ricaurte, "Students' perceptions of generative artificial intelligence (GenAI) use in academic writing in English as a foreign language," *Educ. Sci.*, vol. 15, no. 5, art. no. 611, 2025. https://doi.org/10.3390/educsci15050611.

[4] A. Alshamy, A. Al-Harthi, and S. Abdulla, "Perceptions of generative AI tools in higher education: Insights from students and academics at Sultan Qaboos University," *Educ. Sci.*, vol. 15, no. 4, art. no. 501, 2025. https://doi.org/10.3390/educsci15040501.

[5] S. Farheen, A. Cheema, R. Ullah, and D. Bandeali, "Equity and bias in AI educational tools: A critical examination of algorithmic decision-making in classrooms," *Crit. Rev. Soc. Sci. Stud.*, vol. 3, no. 3, pp. 67-85, 2025. https://doi.org/10.59075/zqmnpa62.

[6] E. Kasneci et al., "ChatGPT for good? On opportunities and challenges of large language models for education," *Learn. Individ. Differ.*, vol. 103, art. no. 102274, 2023. https://doi.org/10.1016/j.lindif.2023.102274.

[7] R. Stockman, "Generative AI and the end of education," in *Proc. Int. Conf. AI Res. (ICAIR)*, vol. 4, no. 1, 2024, pp. 390–397. https://doi.org/10.34190/icair.4.1.3155.

[8] S. Ahmad, A. Ahmed, S. M. Bhutta, and A. N. Ansari, "Transforming teaching learning with chatbots in higher education: Quest, opportunities and challenges for quality enhancement," in *The Evolution of Artificial Intelligence in Higher Education: Challenges, Risks, and Ethical Considerations*, 2024. https://doi.org/10.1108/978-1-83549-486-820241007.

[9] J. Biggs, C. Tang, and G. Kennedy, Teaching for Quality Learning at University, 5th ed. Maidenhead, U.K.: Open Univ. Press, 2023. https://books.google.co.id/books?id=pseVEAAAQBAJ.

[10] F. Eling and A. Ogwal, "Bridging learning gaps: The role of AI-powered technologies in enhancing quality education," *Int. J. Res. Innov. Soc. Sci.*, vol. 9, no. 3, pp. 2011–2016, 2025. https://doi.org/10.47772/IJRISS.2025.903SEDU0155.

[11] J. W. Creswell and V. L. Plano Clark, Designing and Conducting Mixed Methods Research, 3th ed. Thousand Oaks, CA, USA: SAGE, 2017. https://books.google.co.id/books?id=eTwmDwAAQBAJ.

[12] X. Weng, Q. Xia, M. Gu, K. Rajaram, and T. K. Chiu, "Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends," *Australas. J. Educ. Technol.*, vol. 40, no. 6, pp. 37–55, 2024. https://doi.org/10.14742/ajet.9540.

[13] B. Ogunleye, K. I. Zakariyyah, O. Ajao, O. Olayinka, and H. Sharma, "A systematic review of generative AI for teaching and learning practice," *Educ. Sci.*, vol. 14, no. 6, art. no. 636, 2024. https://doi.org/10.3390/educsci14060636.

[14] T. Feng, "ChatGPT's impact on data science students learning performance: A systematic review and prospects," *J. Educ. Educ. Policy Stud.*, vol. 3, no. 1, pp. 1–8, 2024. https://doi.org/10.54254/3049-7248/2024.19048.

[15] L. Gao, "Impact of ChatGPT on academic integrity and assessment effectiveness for e-

learning in higher education: The need for redesigning assessment practices," *Commun. Humanit. Res.*, vol. 45, pp. 40–44, 2024. https://doi.org/10.54254/2753-7064/45/20240084.

[16] K. Žáková, D. Urbano, R. Cruz-Correia, J. L. Guzmán, and J. Matišák, "Exploring student and teacher perspectives on ChatGPT's impact in higher education," *Educ. Inf. Tech.*, vol. 30, pp. 649–692, 2024. https://doi.org/10.1007/s10639-024-13184-y.

[17] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies," *Comput. Educ.*, vol. 227, art. no. 105224, 2025. https://doi.org/10.1016/j.compedu.2024.105224.

[18] J. Roe, M. Perkins, and D. Ruelle, "Understanding student and academic staff perceptions of AI use in assessment and feedback," *arXiv preprint arXiv:2406.15808*, 2024. https://doi.org/10.48550/arXiv.2406.15808.

[19] B. Borges et al., "Could ChatGPT get an engineering degree? Evaluating higher education vulnerability to AI assistants," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 121, no. 49, art. no. e2414955121, 2024. https://doi.org/10.1073/pnas.2414955121.

[20] M. Morrone, "Report: Educators are turning to AI, even for grading," Axios, Aug. 26, 2025. [Online]. Available: https://www.axios.com/2025/08/26/educators-ai-grading-report [Accessed: Oct. 24, 2025].

[21] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quart.*, vol. 13, no. 3, pp. 319–340, Sep. 1989. https://doi.org/10.2307/249008.

[22] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, 2006. https://doi.org/10.1191/1478088706qp063oa.

[23] M. R. Lynn, "Determination and quantification of content validity," *Nurs. Res.*, vol. 35, no. 6, pp. 382–386, 1986. https://doi.org/10.1097/00006199-198611000-00017.