Original Article

Challenges and Strategies of College Teachers in Validating AI-Generated Student Submissions

Jenie Lucero [1,*], Sarrafina Estrebillo [1], Maeanne Caballero [1], Esmeralda Metillo [1]

Received: 10 September 2025; Revised: 15 January 2026;
Accepted: 24 January 2026; Published: 24 January 2026

Abstract

Generative artificial intelligence (AI) tools shape how students draft, revise, and submit academic work. This shift creates a practical dilemma for higher education. Teachers must judge learning evidence when authorship cues weaken and detection tools remain imperfect. This study examined the challenges college teachers face when they validate AI-assisted or AI-generated student submissions and the strategies they apply to protect academic integrity while sustaining meaningful learning. The inquiry used a descriptive qualitative design and drew on semi-structured interviews with ten college teachers from Iligan Medical Center College, selected through purposive sampling, with data collection conducted from February to May 2025. Interview audio was recorded and transcribed, then analyzed through a thematic approach. Results show three main challenge clusters: AI outputs often appear sufficiently fluent to obscure authorship cues; teachers face ethical strain when they suspect AI use without firm proof; and conventional task formats invite over reliance on AI tools, which complicates judgment of students' own thinking. Teachers responded through layered strategies that include transparent rules on acceptable AI use, assessment redesign toward authentic and oral tasks, stronger requirements for citation and disclosure, increased reliance on familiarity with a student's capability, and stricter control of device use during in class tasks.

Keywords: academic integrity, assessment design, generative AI, higher education, thematic analysis

[1] Iligan Medical Center College, Iligan City, Philippines
* Correspondence: jnl36512@imcc.edu.ph

Volume 2, Issue 2, June 2026

1. Introduction

Generative artificial intelligence (AI) has moved from a specialized technology into an everyday academic tool. Text generators can produce coherent explanations, summaries, outlines, and complete essays within seconds, often in a tone that resembles academic prose. For students, this capability can support brainstorming, language refinement, and access to examples. For institutions, however, the same capability can blur the boundary between legitimate assistance and misrepresentation of authorship. Scholars have described generative conversational AI as a system with broad implications for education, research practice, and policy because it changes how people compose text, evaluate knowledge, and distribute expertise (Dwivedi et al., 2023). In education, the concern is not only plagiarism in its traditional form, but also the ease with which a student can submit plausible work without the cognitive work that assessment intends to measure.

Recent literature shows that this dilemma does not have a single technical fix. AI detection tools can produce errors and inconsistent judgments across writing contexts, which makes high stakes decisions difficult. In a controlled evaluation of detection tools, performance varied and false positives remained a concern, which suggests that detection alone cannot serve as a reliable adjudicator of authorship (Elkhatat et al., 2023). As a result, many authors call for assessment and policy responses that align pedagogy, integrity, and fair process rather than sole reliance on software. Commentaries on large language models emphasize that educational benefit depends on deliberate instructional design and clear guidance on ethical use (Alipio, 2024; Kasneci et al., 2023). At the classroom level, educators often face pressure to respond quickly, even while norms remain unsettled.

Higher education also faces a structural challenge: many common assessments depend on unsupervised writing and take home tasks. When AI systems can generate acceptable responses, the evidentiary value of such tasks can weaken unless teachers redesign them to require local context, personal reasoning, or verifiable process. Work on assessment and integrity in the era of ChatGPT argues that academic integrity requires both institutional policy and teaching practices that clarify expectations and reduce incentives for misconduct (Cotton et al., 2024). Related work suggests that debate about the "end" of the traditional essay should prompt serious redesign of tasks toward authenticity, supervision, and demonstration of understanding (Rudolph et al., 2023). Empirical research with educators also reports ambiguity in best practice and highlights assessment modification as a common response (Lindo & Cutad, 2024; Lindo & Panes, 2024).

In the Philippine higher education context, this issue can carry additional constraints tied to resource availability, uneven access to licensed software, variable student digital literacy, and diverse institutional policies. In settings where formal AI governance remains limited, teachers may need to craft informal rules, decide how to confront suspected misuse, and redesign tasks with minimal institutional support. The present study addresses this practical problem by examining teachers' lived classroom

experience in validating AI assisted or AI generated submissions and by documenting strategies that teachers already apply in response. The study focuses on two linked aims: first, to describe challenges teachers encounter when they evaluate student work amid AI generated text; second, to identify strategies teachers use to protect academic integrity while preserving authentic learning.

2. Methodology

This study used a descriptive qualitative approach to examine teachers' experiences and practical responses to AI-generated or AI-assisted student submissions. The setting was Iligan Medical Center College in Iligan City, Northern Mindanao, Philippines. Participants consisted of ten college teachers, with selection guided by purposive sampling and inclusion criteria that required familiarity with AI and the capacity to use it in relation to student outputs. Data collection occurred over three months, from February to May 2025, through semi-structured interviews that used open ended prompts focused on challenges in validation and strategies used in practice. Interviews were conducted after permission from institutional leadership and relevant offices, with sessions held in a setting that supported open discussion; a smartphone captured audio, and recordings were transcribed for analysis. The study applied ethical safeguards that included informed consent, confidentiality protections, and anonymization of identifiers in transcripts.

For analysis, the study used thematic analysis consistent with Braun and Clarke's framework. Researchers read transcripts for familiarity, coded relevant segments, grouped related codes into themes, refined theme definitions, and prepared a narrative report supported by participant excerpts. To strengthen trustworthiness, the analytic process emphasized consistent coding practice, careful handling of discrepant excerpts, and preservation of verbatim quotes to anchor themes in participant accounts.

3. Results

Teachers described challenges that affected both technical judgment and professional ethics when they evaluated suspected AI-assisted work. A central challenge involved AI outputs that appeared sufficiently fluent to hide cues of authorship. Two participants stated that students' AI use had become difficult to detect, with one noting, "*Nowadays the students are better at using AI in a way that's very hard to detect*" (Teacher 5) and another stating, "*It's hard for us teachers to detect whether it's their own ideas or it's from AI*" (Teacher 7).

A second challenge centered on ethical strain and risk of unfair accusation. Three participants highlighted the need for strong evidence prior to confrontation. One teacher stated, "*The teachers really need to be very careful not to say or to accuse the student without strong evidence or proof*" (Teacher 5). Another warned of "*Bias in accusations*" and the possibility of "*damaging trust*" without concrete evidence

(Teacher 6). A third participant described encounters with students who used AI but lacked skill in proper use (Teacher 3).

A third challenge involved assessment formats that teachers viewed as vulnerable to AI dependence and less able to reveal students' own thinking. One participant explained a shift away from seatwork toward tasks that reduce AI reliance: "*I wouldn't give them seat works. I would give them actual assignments*" (Teacher 2). Another emphasized concern that AI could suppress idea formation: "*Due to AI, students may stop developing their own ideas… we want to see their creativity and critical thinking*" (Teacher 7).

Teachers also reported a set of strategies that functioned as layered safeguards rather than a single solution. Some strategies focused on explicit standards and transparency. One participant described setting expectations tied to plagiarism checks: "*If I intend my students to publish their research… they themselves must have to scan their paper for plagiarism*" (Teacher 1). Another emphasized policy clarity: "*Teachers should implement Transparent AI Policies clearly state what AI use is permitted and how it affects their grades*" (Teacher 6).

Many strategies relied on authentic assessment and oral verification to elicit observable understanding. Four participants described approaches such as in class writing, analytic exams, case studies, and oral checks. Examples include "*provide in class writing… teacher can see the students' real abilities in real time*" (Teacher 7), preference for "*case study, analysis… analytical exam*" (Teacher 4), and the use of "*oral exams, presentations… questions and answer sessions to verify understanding*" alongside "*handwritten work*" (Teacher 6).

Other strategies addressed academic conventions and disclosure. Two participants required references even for assignments and permitted AI use only with citation, including "*I also require them to have a reference*" (Teacher 2) and "*I allow them to use AI, but they should cite their reference*" (Teacher 8). Teachers also relied on familiarity with students' typical capability as a practical cue, with one stating that knowing students well was most effective (Teacher 5) and another noting that familiarity made it "*very easy to know if it is AI-generated*" (Teacher 8).

Finally, some participants used technical tools for integrity checks and adopted classroom controls that restrict device access during tasks. One teacher referenced use of Turnitin for plagiarism checks (Teacher 1). Two participants described phone restrictions during in class activity, including collecting phones or banning phones and tablets to reduce reliance on AI tools.

4. Discussion

The findings illustrate a practical integrity crisis that stems from two convergent realities: generative AI can produce fluent academic text, and reliable proof of authorship often remains unavailable in typical classroom conditions. Teachers' description of "hard to detect" outputs aligns with literature that frames

generative AI as a system that alters writing workflows and raises new integrity risks across education (Dwivedi et al., 2023). The challenge is not limited to plagiarism in the conventional copy match sense. When a model produces novel text, similarity tools may not flag it, and detection tools may provide uncertain results. Empirical evaluation of AI content detectors has shown variable efficacy and risk of false positives, which supports teachers' reported lack of confidence in definitive claims (Elkhatat et al., 2023). In this context, teachers' ethical caution reflects not hesitation but professional duty, because erroneous accusation can harm trust, student wellbeing, and the legitimacy of assessment.

Teachers' emphasis on assessment vulnerability and the need to "see" students' thinking also parallels broader calls to reconceptualize assessment. Work on academic integrity in the era of ChatGPT argues that institutions should combine clear expectations with assessment approaches that reduce misconduct incentives and clarify acceptable AI roles (Cotton et al., 2024). The present results show that teachers already enact this principle through authentic tasks, oral verification, and in class writing. Such approaches align with arguments that the traditional essay, especially when unsupervised, no longer serves as a stable proxy for independent thinking and should prompt assessment redesign toward demonstration of understanding (Rudolph et al., 2023). Educator focused research also indicates that many teachers respond through modifications to assessment, often amid uncertainty about best practice, which mirrors the local reports in this study.

The strategy cluster around transparent AI rules, citation requirements, and standards for permitted use reflects a shift from prohibition to governance. This pattern resonates with findings that ethical AI use in education depends on explicit guidance and teacher capacity, not only on technical controls (Kasneci et al., 2023). Systematic review evidence on AI for teachers also highlights a persistent challenge: teachers often assume new roles as evaluators of AI outputs and stewards of responsible classroom integration, which requires training and institutional support (Celik et al., 2022). In practice, teachers in this study relied on familiar indicators of student capability, demanded process evidence such as drafts or oral explanation, and used supervised tasks that reveal reasoning in real time. These strategies function as a triangulation model: the teacher checks alignment between a student's demonstrated understanding, the writing product, and the process evidence.

The reliance on Turnitin and device restrictions illustrates a pragmatic response to limited institutional infrastructure. Yet the literature suggests that overreliance on detection tools can create a false sense of certainty when tool performance varies across prompts, disciplines, and writing styles (Elkhatat et al., 2023). Thus, the most resilient approach appears to combine clear norms, assessment design that elicits authentic performance, and fair process for allegations. Guidance on plagiarism detection in the context of ChatGPT similarly argues for institutional measures that focus on assessment integrity rather than sole dependence on detection, because definitive attribution remains difficult in many cases (Khalil & Er, 2023). The present findings support this stance: teachers did not describe one tool that "solves" the

problem, but rather a set of practices that jointly raise the cost of misconduct and increase the visibility of genuine learning.

Overall, the study indicates that the integrity problem is also an instructional design problem. When tasks privilege generic exposition, AI can fill the gap. When tasks demand situated reasoning, personal justification, and live explanation, teachers gain more valid evidence of learning. This reframes AI from a purely disciplinary threat into a catalyst for stronger assessment validity and clearer academic integrity culture.

5. Conclusion

College teachers in this study faced a core validation dilemma: AI-generated text can resemble student writing, yet proof of misuse often remains uncertain. Teachers reported three interconnected challenge areas: concealment of authorship cues through fluent AI outputs, ethical strain tied to the risk of unfair accusation, and assessment formats that permit AI dependence and weaken evidence of students' own thinking. In response, teachers applied layered strategies that combine transparent expectations for acceptable AI use, stronger citation and disclosure demands, increased use of authentic and oral assessments, reliance on knowledge of student capability, use of integrity tools, and stricter control of device access during in class tasks. These findings suggest that effective institutional response should prioritize assessment validity and fair governance rather than detection alone. Clear AI use policies, teacher development on responsible AI pedagogy, and assessment systems that elicit observable reasoning can support both academic integrity and meaningful learning in a context where generative AI will remain part of student work.

Acknowledgment

Sincere appreciation is given to all peer reviewers for their valuable comments and suggestions, which helped the author to improve the quality of the manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

Alipio, M. M. (2024). Coexist or resist? Impact of artificial intelligence on radiologic technology education. *Journal of Medical Imaging and Radiation Sciences, 55*(4). https://doi.org/10.1016/j.jmir.2024.101450

Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. *Technology in Society, 75*, 102370. https://doi.org/10.1016/j.techsoc.2023.102370

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends, 66*, 616–630. https://doi.org/10.1007/s11528-022-00715-y

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International, 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, K. A., Balakrishnan, J., Barlette, Y., et al. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elkhatat, A. M., Elsaid, K., & Almeer, S. (2023). Evaluating the efficacy of AI content detection tools in differentiating between human and AI generated text. *International Journal for Educational Integrity, 19*(1), Article 17. https://doi.org/10.1007/s40979-023-00140-5

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Geyer, T., Kasneci, G., Krusche, S., & Nerb, J. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. In *Learning and Collaboration Technologies* (Lecture Notes in Computer Science). Springer.

Lindo, M. R., & Cutad, C. A. (2024). Impact of Generative AI on Self-Regulated Learning and Cognitive Offloading. *IMCC Journal of Science, 4*(2), 18-22.

Lindo, M. R., & Panes, R. (2024). The Hallucination Effect: Correlating Generative AI Usage Frequency with Source Verification Habits among Grade 12 STEM Researchers. *IMCC Journal of Science, 4*(2), 9-11.

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching, 6*(1), 342–363. https://doi.org/10.37074/jalt.2023.6.1.9

Author Contributions: Lucero, J., Estrebillo, S., Caballero, M., Metillo, E.; Study design, method conception, data collection, data analysis and manuscript writing