Teacher Monitoring Strategies for AI-Aided Student Writing Outputs Toward Policy Guidelines
DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

Teacher Monitoring Strategies for AI-Aided Student Writing Outputs Toward Policy Guidelines

Masidel M. Bonto
Philippine Christian University
Dasmariñas, Philippines
masidel23@gmail.com

Nora L. Puenteblanca
Philippine Christian University
Dasmariñas, Philippines
nora.puenteblanca@pcu.edu.ph

Abstract

This study examined the extent to which senior high school teachers monitor AI-generated written outputs, focusing on the use of AI detectors, manual review, and cross-checking results. The study respondents were teachers from the Humanities and Social Sciences (HUMSS) strand of Philippine Christian University-Dasmariñas. The researchers collected data through a structured questionnaire and analyzed it using descriptive statistics, including mean and standard deviation. Results showed that teachers used AI detectors and cross-checking to a moderate extent, while they used manual review to a high extent and found it significant. Teachers faced challenges such as limitations in AI detection tools and time constraints in manual evaluation and cross-checking. These findings highlight the importance of a layered monitoring approach, where teacher judgment remains central, and AI tools serve as supplementary aids. The study provides evidence for developing clear guidelines and practices that balance ethical assessment, professional discretion, and responsible integration of AI in student writing evaluation.

***Keywords*:** *Artificial intelligence, AI-generated text, teacher monitoring strategies, manual review, cross-checking*

Bio-profile

Masidel M. Bonto is an educator and researcher affiliated with Philippine Christian University. She earned her Master of Arts in Education, major in English, from Philippine Christian University in 2025 and her Bachelor of Secondary Education, major in English, from the same university in 2022, graduating Magna Cum Laude and consistently earning Dean's List recognition. From 2022 to 2024, she served as a senior high school teacher at Philippine Christian University–Dasmariñas, where she taught creative writing, creative non-fiction, 21st-century literature from the Philippines and the world, and disaster readiness and risk reduction. Her professional interests include English language teaching, student-centered instruction, academic writing, artificial intelligence in education, academic integrity, and teacher monitoring of AI-aided writing. Beyond her teaching responsibilities, she has also participated in university theatre and choir activities, dedicating her musical talents to glorifying the Lord through worship.

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

Introduction

Generative artificial intelligence has changed how students produce and revise academic writing. Tools that generate, paraphrase, edit, and provide immediate feedback can support fluency and accuracy, but their increasing use also raises concerns about authorship, originality, academic integrity, and the validity of writing assessment. Recent literature describes both the educational potential of generative artificial intelligence and the risks that arise when its use is insufficiently guided or regulated (Bjelobaba et al., 2024; Akbar, 2025).

The issue is particularly relevant to writing-intensive senior high school contexts, where teachers must evaluate not only the quality of a submitted text but also the authenticity of the student's learning process. Philippine studies cited in the source report indicate growing adoption of generative artificial intelligence alongside uneven institutional readiness, teacher preparation, and policy implementation (Fudalan, 2025; Alvero & Saldua, 2026).

Although research has examined the accuracy of artificial intelligence detection tools, it has paid less attention to how teachers combine detection, direct manual review, and cross-checking of evidence in everyday assessment. Detection systems can produce false classifications and become less reliable when students paraphrase or edit generated text (Elkhatat et al., 2023; Khalil & Er, 2023; Liang et al., 2023). Human review can identify contextual and stylistic cues but depends on teacher expertise and takes time (Waltzer et al., 2024; Simmons & Ward, 2025). Cross-checking and triangulation have been proposed as ways to strengthen judgments, yet their practical use remains comparatively underexplored (Benarab, 2024).

This study addressed this gap by examining the monitoring strategies used by senior high school humanities and social sciences teachers at Philippine Christian University–Dasmariñas. It focused on three strategies—AI detectors, manual review, and cross-checking results—across word choice (vocabulary use), structure and style, and error patterns. It also identified challenges teachers encountered and policy needs that could support clear, consistent monitoring of AI-aided student writing. The study therefore aimed to provide an institutionally relevant basis for ethical, fair, and practical monitoring guidelines.

Theoretical Framework

The study is anchored in constructivist learning theory, the technology acceptance model, and academic integrity theory. Together, these perspectives frame monitoring as a pedagogical, technological, and ethical process.

Constructivist learning theory emphasizes active and reflective learning. In writing instruction, the teacher's role is not merely to judge the final product but to guide students in constructing knowledge through drafting, revision, reflection, and meaningful engagement (Piaget, 1952; Harasim, 2018). Manual review and verification of prior work are consistent with this view because they keep attention on the student's learning process and individual development.

The Technology Acceptance Model explains technology adoption through perceived

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

usefulness and perceived ease of use (Davis, 1989; Marangunić & Granić, 2015). In this study, teachers' moderate use of AI detectors and cross-checking may be understood in relation to their perceptions of these tools' usefulness, limitations, reliability, and practicality.

Academic Integrity Theory emphasizes honesty, fairness, responsibility, and contextual influences on academic misconduct (McCabe & Treviño, 1993; Bretag, 2016). This perspective supports policies that distinguish responsible AI assistance from misrepresentation while protecting students and teachers from decisions based solely on uncertain detection evidence.

Conceptual Framework

This framework presents the monitoring strategies for AI-aided writing output using AI detectors, manual review, and cross-check results, leading to the establishment of a clear institutional policy on monitoring AI-aided writing output.

Figure 1. Conceptual Framework

The conceptual framework illustrates the monitoring strategies for AI-aided writing outputs as a multidimensional process leading toward the establishment of a clear institutional policy. The monitoring strategies were categorized into three primary domains: AI Detectors, Manual Review, and Cross-Check Results.

Across three indicators—word choice/vocabulary use, structure and style, and error patterns—teachers applied each strategy. The effectiveness and consistency of these strategies influence the teachers' ability to identify AI-generated content accurately. However, challenges such as limited training, tool inaccuracy, insufficient school policies, and time constraints may hinder the monitoring process. Understanding these strategies and challenges provided the basis for developing clear policies on monitoring strategies for AI-generated texts, ensuring consistency, fairness, and ethical integrity in assessing AI-aided student writing outputs.

Research Questions

The study sought to answer the following questions:

1. To what extent do teachers currently use monitoring strategies to assess AI-generated texts, including AI detectors, manual review, and cross-checking results across word choice/vocabulary, structure and style, and error patterns?
2. What challenges do teachers encounter in implementing these monitoring strategies?
3. Based on the findings, what policy guidelines can be developed to establish clear and consistent monitoring strategies for AI-generated texts?

Literature Review

AI Detector Monitoring Strategies. A growing body of research has critically evaluated the reliability and validity of AI-detection tools used to identify machine-generated text. Elkhatat, Elsaid, and Almeer (2023) reported that across fifteen paragraphs comprising both ChatGPT-generated and human-written samples, detection software achieved only moderate success and frequently produced

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

false classifications. Similarly, a comprehensive evaluation involving twelve publicly available detectors and two commercial systems revealed major reliability concerns. The study concluded that existing tools "are neither accurate nor reliable and have a main bias towards classifying the output as human-written rather than detecting AI-generated text" (Elkhatat et al., 2023; Waltzer et al., 2024). Further research shows that these tools' effectiveness declines substantially when AI-generated outputs are paraphrased, revised, or repeatedly edited, making them increasingly difficult to distinguish from human-written text (Khalil & Er, 2023; Liang et al., 2023). These findings suggest that relying solely on a single AI detector is insufficient; instead, combining multiple detection tools with human judgment can improve the reliability of authenticity assessments. In the medical writing domain, one study found that although detectors such as Originality.ai and ZeroGPT performed reasonably well on unedited AI-generated texts, their accuracy decreased considerably after stylistic revisions and paraphrasing (Elkhatat et al., 2023).

Manual Review and Human Oversight. Manual review by teachers remains central to verifying the authenticity of student writing despite technological advances. Human review involves assessing subtle linguistic and contextual cues, such as repetitive sentence patterns, absence of personal voice, or overly formal phrasing, that often reveal AI influence. Studies indicate that teachers can identify AI-generated essays with moderate accuracy; one investigation found a 68% correct detection rate, albeit with a 14% false-positive rate (Liang et al., 2024). However, reviewers' accuracy varies significantly by confidence, training, and teaching experience. Waltzer, Pilegard, and Heyman (2024) further established that while instructor expertise and self-assuredness predicted stronger detection ability, considerable variance persisted even among experienced educators (Elkhatat et al., 2023).

Cross-Check and Triangulation. Cross-checking—using multiple sources of evidence such as detector results, manual evaluation, and document draft histories—is an emerging topic in the literature. Scholars argue that triangulation enhances reliability by reducing both false positives and negatives through a multi-layered verification process. Benarab (2024), for example, demonstrated that employing several AI-detection tools concurrently improved accuracy in identifying AI-generated text, emphasizing "the need to combine several to overcome their deficiencies" (Benarab, 2024). Despite its potential, triangulated monitoring remains underexplored in educational research, particularly within secondary and higher education institutions in developing contexts.

Policy, Institutional Frameworks, and Ethics. Beyond the mechanics of detection, scholars emphasize the importance of coherent institutional policies governing AI use and monitoring. Studies consistently highlight fragmented and inconsistent approaches across educational systems. Liang (2024) observed that many institutions lack standardized guidelines or sufficient awareness regarding AI utilization in academic writing. Simmons and Ward (2025) found that students' ethical beliefs and values were stronger predictors of AI use than policy presence alone, suggesting that institutional rules must be paired with training and culture-building initiatives to be effective. Ethical education and clear institutional communication are thus critical in shaping responsible AI practices among both teachers and students.

Learning and Assessment Implications. The debate on AI monitoring extends beyond detection into broader pedagogical concerns. Overreliance on AI-generated content risks bypassing essential stages of learning such as idea development, reflection, and critical analysis. Scholars have argued that instead of enforcing outright bans, educators should explore "AI-inclusive" assessment

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

models that integrate technology while promoting cognitive engagement. Perkins et al. (2023), in a study assessing GPT-4 outputs against student work, found comparable performance levels between the two, suggesting that assessment design, not detection alone, should evolve to uphold learning integrity (Rafiq, Qurat-ul-Ain, & Afzal, 2025). Redesigning writing tasks to emphasize process-based evaluation, such as drafts, peer feedback, and reflective commentary, may help to maintain authentic learning even as AI becomes more widely used.

Overall, current research highlights several unresolved issues in monitoring AI use in education. The results of current tools, like AI detectors and manual reviews, are often inconsistent and lack standardized processes, and cross-checking techniques remain rarely investigated, particularly in developing nations and schools. Institutional policies on AI usage are also uneven or lacking, with limited attention to ethical and cultural aspects. Very few studies have combined these monitoring strategies into a unified, context-based framework, making it essential to understand how teachers currently apply them and to design strategies that promote fairness and integrity in student writing in the age of generative AI.

Synthesis

The reviewed studies reveal that generative AI tools have reshaped the landscape of academic writing, creating both opportunities and ethical challenges for educators and students. Various monitoring strategies, AI detectors, manual review, and cross-checking have been explored to ensure writing authenticity (Elkhatat, Elsaid, & Almeer, 2023; Benarab, 2024; Waltzer, Pilegard, & Heyman, 2024). However, these strategies face recurring reliability and subjectivity issues. AI detectors often misclassify text and perform inconsistently when outputs are paraphrased or edited (Khalil & Er, 2023; Liang et al., 2023; Elkhatat et al., 2023), while manual reviews depend heavily on teacher expertise and confidence (Waltzer et al., 2024; Simmons & Ward, 2025). Triangulation, or combining multiple methods, has been recommended to improve accuracy, yet remains underexamined in many educational contexts, especially in developing countries where resources and training are limited (Benarab, 2024).

Moreover, policy and ethics literature emphasizes fragmented institutional approaches to AI governance. Although some universities have issued responsible AI-use guidelines, many schools, particularly in the Philippines, lack structured frameworks that support teachers in monitoring AI-assisted writing (Liang, 2024). Scholars argue that integrity cannot rely solely on detection or policy but must be reinforced through ethical awareness and redesigned assessments that prioritize process, reflection, and critical thinking (Perkins et al., 2023; Rafiq, Qurat-ul-Ain, & Afzal, 2025). Thus, the gap lies in the absence of an integrated, context-sensitive monitoring framework that combines detection tools, human judgment, and cross-verification to uphold academic integrity in AI-assisted learning environments.

Method

Research Design

The study used a quantitative descriptive design to describe teachers' monitoring strategies and examine whether item-level mean ratings differed from a test value of 3.5. This design was appropriate for summarizing numerical responses and applying inferential analysis to the monitoring indicators.

Participants

The participants were 41 senior high school teachers from the Humanities and Social Sciences

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

strand at Philippine Christian University–Dasmariñas during School Year 2025–2026. The researcher used convenience sampling because the participants were accessible. The findings are therefore specific to the participating teachers and institution and should not be generalized to other schools or academic programs without further evidence.

Instrument

Data were collected using a researcher-made questionnaire with sections on AI detectors, manual review, cross-checking results, and challenges encountered in implementing monitoring mechanisms. Items addressed word choice or vocabulary use, structure and style, and error patterns. The questionnaire was validated and pilot-tested with five teachers who were not included in the actual sample. The reported Cronbach's alpha was 0.95, indicating strong internal consistency.

Ethical Considerations

Participation was voluntary, and the researchers obtained informed consent before data collection. The study maintained confidentiality and anonymity and used responses solely for academic and research purposes. The school administration granted permission to conduct the study. These procedures were applied throughout questionnaire distribution, retrieval, data preparation, and analysis.

Statistical Treatment

Means and standard deviations described the extent of monitoring practices. The scale used five levels: 4.50–5.00, Very High Extent; 3.50–4.49, High Extent; 2.50–3.49, Moderate Extent; 1.50–2.49, Low Extent; and 1.00–1.49, Very Low Extent. One-sample t-tests used a test value of 3.5, with $\alpha = .05$ and $df = 8$ for the item-level analyses. Frequencies and percentages summarized challenges and identified policy needs.

Results and Discussion

1. Extent of the Current Monitoring Strategies Used by the Teacher in Evaluating AI-Aided Writing Output

Table 1

Extent of Teachers' Use of AI Detectors

Indicators	Mean	SD	Interpretation
Word choice/vocabulary use			
1. Identification of repetitive ideas	2.73	1.27	Moderate Extent
2. Detection of repetitive or generic word choices	2.88	1.31	Moderate Extent
3. Identification of unnecessary or redundant words	2.61	1.09	Moderate Extent
Structure and Style			
4. Detection of unusually uniform sentence structures	2.78	1.08	Moderate Extent
5. Identification of writing patterns lacking personal voice	2.63	1.09	Moderate Extent
6. Detection of inconsistencies in writing style	2.59	1.05	Moderate Extent
Error Patterns			
7. Detection of inconsistencies in error patterns	2.51	1.08	Moderate Extent
8. Flagging polished grammar with weak logical development	2.80	1.17	Moderate Extent

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

9. Identification of inconsistencies with students' usual errors	2.88	1.29	Moderate Extent
Overall Mean	**2.71**	—	**Moderate Extent**

Table 1 shows that AI detectors were used to a moderate extent (M = 2.71). The highest means were for repetitive or generic word choices and inconsistencies with students' usual errors (both M = 2.88), while inconsistencies in error patterns had the lowest mean (M = 2.51). The pattern suggests selective rather than routine use. This is consistent with literature showing that detector accuracy varies and can decline after paraphrasing or stylistic revision (Elkhatat et al., 2023; Khalil & Er, 2023; Liang et al., 2023).

Table 2
Extent of Teachers' Use of Manual Review

Indicators	Mean	SD	Interpretation
Word choice/vocabulary use			
1. Manual identification of repetitive ideas	3.73	1.07	High Extent
2. Evaluation of repetitive or generic word choices	3.63	1.11	High Extent
3. Identification of unnecessary or redundant words	3.59	1.12	High Extent
Structure and Style			
4. Assessment of unusually uniform sentence structures	3.59	1.18	High Extent
5. Evaluation of writing lacking personal voice	3.39	1.14	Moderate Extent
6. Examination of inconsistencies in writing style	3.46	1.12	Moderate Extent
Error Patterns			
7. Checking inconsistencies in error patterns	3.37	1.16	Moderate Extent
8. Identification of polished grammar with weak logic	3.56	1.14	High Extent
9. Analysis of inconsistencies with students' usual errors	3.56	1.14	High Extent
Overall Mean	**3.53**	—	**High Extent**

Table 2 shows that manual review obtained the highest overall mean (M = 3.53), indicating High Extent. Teachers most frequently reported manually identifying repetitive ideas (M = 3.73). However, evaluation of personal voice, writing-style inconsistencies, and error-pattern inconsistencies were only moderate. The finding reinforces the importance of teacher expertise and contextual knowledge in authenticity assessment, while also showing that nuanced cues may be more difficult to apply consistently (Waltzer et al., 2024).

Table 3
Extent of Teachers' Use of Cross-Checking Results

Indicators	Mean	SD	Interpretation
Word choice/Vocabulary use			
1. Cross-checking repetitive ideas	3.22	1.11	Moderate Extent
2. Verification of vocabulary use	3.00	1.07	Moderate Extent
3. Validation of redundant word usage	3.02	1.06	Moderate Extent
Structure and Style			
4. Comparison of sentence structure assessments	2.98	1.01	Moderate Extent

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

5. Verification of personal voice using prior writing	3.12	1.08	Moderate Extent
6. Validation of stylistic inconsistencies	3.20	1.05	Moderate Extent
Error Patterns			
7. Cross-checking error pattern inconsistencies	3.10	1.02	Moderate Extent
8. Verification of grammar-quality mismatches	3.00	1.02	Moderate Extent
9. Comparison with previous writing samples	3.12	1.00	Moderate Extent
Overall Mean	**3.08**	—	**Moderate Extent**

Table 3 shows that teachers used cross-checking to a moderate extent (M = 3.08), with all indicators falling within the moderate range. Cross-checking repetitive ideas had the highest mean (M = 3.22), while comparing sentence-structure assessments had the lowest (M = 2.98). This indicates that teachers use corroborative evidence selectively rather than through a standardized routine. The finding supports the need for institutional procedures that make cross-checking practical and consistent (Benarab, 2024).

Overall, the results show a clear difference in how much teachers used the three monitoring strategies. Manual review obtained the highest overall mean and was classified as High Extent, whereas AI detectors and cross-checking were both classified as Moderate Extent. This pattern indicates that teachers continue to place greater weight on direct professional judgment while using technology and cross-verification as supporting mechanisms.

2. Challenges Encountered by Teachers

Table 4
Challenges that Teachers Encounter in Implementing Monitoring Mechanisms

Challenges	Frequency	Percentage
1. Unreliability of AI detection tools	29	70.73%
2. Time constraints for manual review and cross-checking	26	63.41%
3. Lack of training in AI monitoring tools	23	56.10%
4. Difficulty distinguishing AI-assisted from genuine student improvement	20	48.78%
5. Absence of clear institutional policies on AI use	18	43.90%

Table 4 shows that the most prevalent challenge was the unreliability of AI detection tools (70.73%), followed by time constraints for manual review and cross-checking (63.41%) and lack of training in AI monitoring tools (56.10%). Teachers also reported difficulty distinguishing AI assistance from genuine student improvement (48.78%) and the absence of clear institutional policies (43.90%). These results show that effective monitoring is constrained not only by technology but also by workload, professional preparation, and institutional clarity.

3. Implications for Policy Guideline Development

Table 5

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

Identified Needs for Policy Guideline Development in Monitoring AI-aided Student Writing Output

Identified Policy Needs	Frequency	Percentage
1. Formal training on AI literacy and AI detection tools	31	75.61%
2. Clear institutional policies defining acceptable AI use	28	68.29%
3. Standardized monitoring procedures combining AI tools and manual review	26	63.41%

Shown in Table 5, formal training was the leading policy need (75.61%), followed by clear institutional definitions of acceptable AI use (68.29%) and standardized procedures combining AI tools with manual review (63.41%). These findings point to system-level support rather than reliance on individual teacher judgment alone. Policy development should therefore combine teacher training, clear student expectations, workable procedures, and evidence-based decision rules.

One-Sample t-Test Findings

The one-sample t-tests were conducted for each indicator using a test value of 3.5 at the 0.05 level of significance (df = 8). A Sig. (2-tailed) value below .05 indicates a statistically significant difference from the test value. The complete item-level results for each monitoring strategy are presented below.

Table 6

One-Sample t-Test Result on the Extent of Teachers' Use of AI-Detectors

Indicator	Test Value=3.5						
	Mean	df	t	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
						Lower	Upper
Word choice/Vocabulary use							
1. Identification of repetitive ideas	2.73		-1.82	0.106	-0.77	-1.74	0.2
2. Detection of repetitive or generic word choices	2.88	8	-1.43	0.191	-0.62	-1.62	0.38
3. Identification of unnecessary or redundant words	2.61		-2.45	0.04	-0.89	-1.72	-0.06
Structure and Style							
4. Detection of unusually uniform sentence structures	2.78		-2.00	0.08	-0.72	-1.55	0.11
5. Identification of writing patterns lacking personal voice	2.63		-2.39	0.044	-0.87	-1.7	-0.04
6. Detection of inconsistencies in writing style	2.59	8	-2.60	0.031	-0.91	-1.71	-0.11
Error Patterns							

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

7. Detection of inconsistencies in error patterns	2.51		-2.75	0.025	-0.99	-1.82	-0.16
8. Flagging polished grammar with weak logical development	2.80	8	-1.79	0.112	-0.7	-1.6	0.2
9. Identification of inconsistencies with students' usual errors	2.88		-1.44	0.188	-0.62	-1.61	0.37

Table 6 shows that four AI-detector indicators significantly differed from the test value of 3.5: identification of unnecessary or redundant words (M = 2.61, t = -2.45, p = .040), writing patterns lacking personal voice (M = 2.63, t = -2.39, p = .044), inconsistencies in writing style (M = 2.59, t = -2.60, p = .031), and inconsistencies in error patterns (M = 2.51, t = -2.75, p = .025). The remaining five indicators were not statistically significant: repetitive ideas (p = .106), repetitive or generic word choices (p = .191), unusually uniform sentence structures (p = .080), polished grammar with weak logical development (p = .112), and inconsistencies with students' usual errors (p = .188).

Table 7

One-Sample t-Test Result on the Extent of of Teachers' Use of Manual Review

Indicator	Test Value=3.5						
	Mean	**df**	**t**	**Sig. (2-tailed)**	**Mean Difference**	**95% Confidence Interval of the Difference**	
						Lower	**Upper**
Word choice/Vocabulary use							
1. Manual identification of repetitive ideas	3.73		0.64	0.54	0.23	-0.59	1.05
2. Evaluation of repetitive or generic word choices	3.63		0.35	0.74	0.13	-0.72	0.98
3. Identification of unnecessary or redundant words	3.59		0.24	0.82	0.09	-0.77	0.95
Structure and Style							
4. Assessment of unusually uniform sentence structures	3.59		0.23	0.83	0.09	-0.82	1
5. Evaluation of writing lacking personal voice	3.39		-0.29	0.78	-0.11	-0.99	0.77
6.Examination of inconsistencies in writing style	3.46		-0.11	0.91	-0.04	-0.9	0.82
Error Patterns							

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

7. Checking inconsistencies in error patterns	3.37	-0.34	0.74	-0.13	-1.02	0.76
8. Identification of polished grammar with weak logic	3.56	0.16	0.88	0.06	-0.82	0.94
9. Analysis of inconsistencies with students' usual errors	3.56	0.16	0.88	0.06	-0.82	0.94

Table 7 presents the complete manual-review results. None of the nine indicators significantly differed from the test value of 3.5. The p-values ranged from .540 to .910, indicating no statistically significant difference for any manual-review indicator at the 0.05 level. This reinforces the role of manual review as a regularly practiced and reliable method in evaluating AI-assisted student writing.

Table 8

One-Sample t-Test Result on the Extent Teachers' Use of Cross-Checking Results

Indicator	Test Value=3.5					
	Mean	df	t	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference
						Lower / Upper
Word choice/Vocabulary use						
1. Cross-checking repetitive ideas	3.22		-0.76	0.469	-0.28	-1.13 / 0.57
2. Verification of vocabulary use	3.00		-1.4	0.199	-0.5	-1.32 / 0.32
3. Validation of redundant word usage	3.02		-1.36	0.21	-0.48	-1.29 / 0.33
Structure and Style						
4. Comparison of sentence structure assessments	2.98		-1.54	0.162	-0.52	-1.29 / 0.25
5. Verification of personal voice using prior writing	3.12		-1.06	0.321	-0.38	-1.21 / 0.45
6. Validation of stylistic inconsistencies	3.20		-0.86	0.414	-0.3	-1.11 / 0.51
Error Patterns						
7. Cross-checking error pattern inconsistencies	3.10		-1.18	0.27	-0.4	-1.18 / 0.38
8. Verification of grammar-quality mismatches	3.00		-1.47	0.18	-0.5	-1.28 / 0.28
9. Comparison with previous writing samples	3.12		-1.14	0.288	-0.38	-1.15 / 0.39

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

Table.8 presents the complete cross-checking results. None of the nine indicators significantly differed from the test value of 3.5. The p-values ranged from .162 to .469, indicating no statistically significant difference for any cross-checking indicator at the 0.05 level.

Four AI-detector indicators significantly differed from 3.5: unnecessary or redundant words ($p = .040$), lack of personal voice ($p = .044$), writing-style inconsistencies ($p = .031$), and error-pattern inconsistencies ($p = .025$). Their mean differences were negative, indicating ratings below 3.5. The remaining AI-detector indicators were not significant. All manual-review and cross-checking indicators were also not significant. Thus, manual review had the highest descriptive level of use, but its indicators did not significantly differ from the test value.

The findings support a layered monitoring procedure. Manual review can provide the primary contextual assessment, while AI detectors can serve as supplementary screening tools and cross-checking can corroborate concerns through drafts, prior writing, teacher observations, and other available evidence. Decisions should be based on multiple forms of evidence and documented procedures.

Policy development should also address transparency and student agency. Students should receive clear guidance on acceptable and prohibited AI use and disclosure expectations. Teachers need continuing AI-literacy training, while administrators should provide practical procedures, adequate assessment time, and regular policy review.

Conclusions

1. Teachers used the manual to the greatest extent, while AI detectors and cross-checking were used to a moderate extent in monitoring student writing outputs. This indicates that teacher judgment remains the principal practical monitoring approach.
2. The findings indicate that AI detectors, manual review, and cross-checking should not be used independently, as no single strategy is sufficient to establish the authenticity of student writing.
3. Teachers' monitoring practices are constrained by unreliable detection tools, time demands, limited training, difficulty distinguishing AI assistance from genuine improvement, and unclear institutional policy. Key policy needs include formal AI literacy training, clear definitions of acceptable AI use, and standardized procedures that integrate technology with human judgment.
4. An effective institutional approach should therefore be layered, transparent, evidence-based, and ethically grounded, with manual review at the center and AI detection and cross-checking used as supporting evidence rather than as stand-alone grounds for disciplinary decisions.

Recommendations

Based on the conclusions, the following are recommended:

1. Teachers should receive ongoing training in AI literacy and monitoring strategies. Professional development programs should focus on understanding AI-generated writing patterns, ethical AI use, and how to interpret AI detector results properly. This will strengthen teachers' confidence, improve evaluation accuracy, and reduce overreliance on or mistrust of AI tools.
2. Teachers should continue prioritizing manual review as the primary monitoring strategy, supported by AI tools. Because manual review was statistically significant, teachers should rely on their professional judgment and familiarity with students' writing styles while using AI detectors and cross-checking as supplementary tools.

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

3. Students should be oriented to be responsible and ethical in the use of AI in academic writing. Communicate clear guidelines to help students understand acceptable AI assistance versus academic dishonesty. This promotes transparency, academic integrity, and responsible AI use, while reducing confusion and unintentional violations.
4. Students should be encouraged to develop authentic writing skills alongside AI use. Instructional activities should emphasize critical thinking, personal voice, and originality to ensure that AI serves as a learning aid rather than a replacement for genuine student effort.
5. School administrators should establish clear, standardized institutional policies on AI use in student writing. These policies should define acceptable AI practices, monitoring procedures, and consequences for misuse. Clear policies will ensure consistency, protect teachers from subjective decision-making, and promote fairness in student assessment.
6. Administrators should design realistic monitoring frameworks that consider teachers' workload. Given the reported time constraints, administrators should support streamlined monitoring procedures and provide institutional support, such as access to tools, reference materials, and reduced administrative burden.
7. Future studies should examine the long-term effectiveness of integrated AI monitoring strategies. Researchers may explore longitudinal impacts of teacher training, differences across academic strands, or comparative effectiveness of various AI detection tools.
8. Further research may focus on students' perspectives and learning outcomes related to AI use. Investigating how AI monitoring policies affect student motivation, writing development, and ethical awareness will provide a more holistic understanding of AI integration in education.
9. Schools should adopt and implement the proposed AI Monitoring Policy Guidelines developed in this study to guide monitoring of AI-assisted student writing. The framework may serve as the basis for developing institutional policies, standard operating procedures, and teacher training programs on the ethical and responsible use of generative AI in academic writing. Its implementation is expected to promote consistency in monitoring practices, strengthen academic integrity, support teachers in making informed evaluation decisions, and give students clear expectations about acceptable AI use in educational settings.

Acknowledgment

The author sincerely expresses her gratitude to all individuals and institutions whose guidance, support, and contributions made this study possible.

She is deeply grateful to her adviser and panel members for their valuable guidance, constructive feedback, and recommendations, which greatly contributed to improving this research. She also extends her appreciation to the administration of **Philippine Christian University–Dasmariñas** for granting permission to conduct the study and to the senior high school humanities and social sciences teachers who participated in the research and cooperated during the data-gathering process.

The author likewise thanks her friends for their encouragement and moral support throughout the academic journey. Above all, she is profoundly grateful to her family for their unwavering love, support, and sacrifices, and to God for the strength, wisdom, and guidance that sustained her throughout this study.

References

Akbar, M. (2025). Beyond detection: Designing AI-resilient assessments with automated

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

feedback tools to foster critical thinking. arXiv. https://arxiv.org/abs/2503.23622

Alvero, J. C. M., & Saldua, M. P. (2026). Teachers' perceptions and utilization of generative AI in assessment and feedback: Evidence for policy development in basic education. Journal of Interdisciplinary Perspectives, 4(6). https://ejournals.ph/article.php?id=34633

Benarab, I. H. (2024). Detection of AI-generated writing in students' assignments: A comparative analysis of some tools' reliability. ATRAS Journal. https://doi.org/10.70091/atras/AI.17

Bjelobaba, S., Wang, S., & Petrova, N. (2024). The influence of generative AI on academic authorship and integrity. Journal of Academic Ethics, 22(3), 125–142.

Bretag, T. (2016). Defining academic integrity: International perspectives. Journal of Higher Education Policy and Management, 38(3), 316–329.

Crompton, H., Burke, D., Bozkurt, A., et al. (2026). A framework for generative AI policy and guidelines in K–12 education. Journal of Research on Technology in Education. https://doi.org/10.1080/15391523.2026.2661641

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.

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, 17. https://doi.org/10.1007/s40979-023-00140-5

Fudalan, A. (2025). Can AI replace teachers? The value of artificial intelligence in Philippine education. Psychology and Education: A Multidisciplinary Journal, 43(4). https://ejournals.ph/article.php?id=30510

Harasim, L. (2018). Constructivist learning theory. In Learning Theory and Online Technologies (pp. 61–79).

Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking plagiarism detection. arXiv. https://doi.org/10.48550/arXiv.2302.04335

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. Patterns, 4(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Marangunić, N., & Granić, A. (2015). Technology acceptance model: A literature review. Information Systems Frontiers, 17(3), 517–551.

McCabe, D. L., & Treviño, L. K. (1993). Academic dishonesty: Honor codes and other contextual influences. The Journal of Higher Education, 64(5), 522–538. https://doi.org/10.1080/00221546.1993.11778446

Piaget, J. (1952). The origins of intelligence in children. International Universities Press.

Simmons, S., & Ward, E. (2025). Student perceptions of AI-assisted writing and academic integrity: Ethical concerns, academic misconduct, and use of generative AI in higher education. AI in Education, 1(1), 2. https://doi.org/10.3390/aieduc1010002

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53–55.

This work is licensed under a Creative Commons Attribution 4.0 International License.

DOI: https://doi.org/10.56738/issn29603986.geo2026.7.247
ISSN 2960-3986
GEO Academic Journal Vol. 7 No. 4

Waltzer, T., Pilegard, C., & Heyman, G. D. (2024). Can you spot the bot? Identifying AI-generated writing in college essays. International Journal for Educational Integrity, 20, 11. https://doi.org/10.1007/s40979-024-00158-3

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J. G., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19(1), 1–39. https://doi.org/10.1007/s40979-023-00146-z

This work is licensed under a Creative Commons Attribution 4.0 International License.