

Generative AI and Academic Integrity in Online and Distance Learning: A Policy Index and Evidence for Assessment Redesign in the Global South

Sixbert Sangwa*¹, Placide Mutabazi²

ABSTRACT

Generative AI has unsettled assessment and integrity practice, especially for large-scale ODL institutions in the Global South. Using a scoping meta-synthesis of policy texts ($n = 50$) and public indicators for 40 ODL providers, we propose and apply the AI-aware Assessment Policy Index (AAPI), positioning institutions along redesign-versus-surveillance and transparency-versus-opacity axes. Higher AAPI scores correlate with lower AI-detector flag rates and reduced proctoring reliance (Spearman $\rho = -0.42$, $p < .01$), without evidence of increased misconduct. We contextualize these findings within a sociotechnical, value-sensitive model, noting persistent risks from biased AI-text detectors and error-prone facial analytics that overburden multilingual and low-bandwidth learners. Who benefits. The AI-aware Assessment Policy Index (AAPI) is intended for ODL leaders, program heads, assessment designers, quality-assurance officers, and national regulators seeking evidence-informed alternatives to blanket detection and high-surveillance regimes. Methods transparency. We scored six policy dimensions (assessment redesign, disclosure norms, AI literacy, process safeguards, equity safeguards, and surveillance intensity), normalized each to $[0,1]$, and computed the AAPI as the simple average of the six dimensions; inter-coder reliability exceeded the .80 benchmark across all dimensions. We also report internal consistency of the six-dimension scale (Cronbach's α) and provide a per-institution score table in the Supplement. The contribution is a concise, adaptable blueprint: re-design assessments, codify disclosure, invest in AI literacy, and embed due-process and equity safeguards. We conclude that credible integrity in the AI era is a design outcome, not a surveillance outcome, particularly in resource-constrained ODL systems. Given the deliberately modest sample ($n = 40$ institutions), the study should be read as a pilot that seeds a new research paradigm rather than offering definitive causal claims. Its value lies in furnishing a transferable blueprint whose propositions invite testing across larger, multi-regional datasets.

Keywords: generative AI in education; academic integrity; assessment redesign; online and distance learning; Global South; policy innovation; AI governance

Submitted: September 30, 2025

Accepted : December 10, 2025

Published : December 11, 2025

Authors' Affiliation

¹Department of International Business and Trade, African Leadership University, Kigali, Rwanda.

²Office of the Executive Chancellor, Open Christian University, California, the United States

*Corresponding Author

Sangwa, Sixbert: ssangwa@olueducation.com

Cite this as:

Sangwa, S. & Mutabazi P. (2025). *Generative AI and Academic Integrity in Online and Distance Learning: A Policy Index and Evidence for Assessment Redesign in the Global South*. Open Journal of AI, Ethics and Society, Vol 1(2).

DOI: [10.5281/zenodo.17902153](https://doi.org/10.5281/zenodo.17902153)

1. Introduction

Universities entered the generative-AI moment with fragile post-pandemic infrastructures and unresolved equity gaps, and ODL providers in the Global South felt these pressures most acutely. Early reactions were blunt: moratoria on AI tools and rapid adoption of AI-text detectors and remote proctoring. These moves proved brittle as evidence of detector bias and student backlash accumulated, especially among multilingual learners and those with limited connectivity (Liang, Yuksekgonul, Mao, Wu, & Zou, 2023; Human Rights Watch, 2022). A more durable posture is now visible in global guidance and national quality-assurance discourse: move beyond bans, redesign assessment, and embed disclosure, literacy, and due-process protections (UNESCO, 2023; TEQSA, 2023; OECD, 2023). Building on these currents, we synthesize institutional policy responses and test whether design-oriented policies align with healthier integrity signals in ODL contexts. We offer an indexable model and a concise

blueprint for leaders seeking credibility without escalating surveillance. ODL students in the Global South face additional inequities – from patchy internet access to linguistic mismatches with mainstream AI tools (see Sangwa & Mutabazi, 2025b) – raising the stakes for ethical AI governance that is sensitive to low-resource contexts (Jin et al., 2024; North, 2023; UNICEF, 2020, 2021). Accordingly, the index and blueprint are designed for ODL executives, QA agencies, and regulators who must balance integrity, equity, and feasibility under tight bandwidth and capacity constraints (APRU, 2025; EDUCAUSE, 2025).

Since early 2025, sector evidence has underscored both the velocity of student uptake and the limits of detector-led responses. A UK-wide survey of 1,041 undergraduates found that 92% now use some form of AI and 88% have applied GenAI in assessment-adjacent ways, with institutions urged to 'stress-test' assessments rather than rely on detection alone (Higher Education Policy Institute [HEPI], 2025). Comparative reviews of university

guidance show rapid formalization of GenAI policies across leading institutions, with emphasis on assessment redesign, disclosure, and role-specific guidance (An, Yu, & James, 2025). System overviews likewise document education as a major locus of AI adoption and policy experimentation in 2025 (Stanford HAI, 2025; Ilieva et al., 2025). These developments reinforce our focus on design-first integrity for ODL.

1.1. Problem Statement and Gap: In the rush to respond to AI, universities have moved quickly, but not coherently. Many ODL institutions initially doubled down on technocratic measures – adopting AI text detectors or expanding online proctoring – despite mounting evidence of reliability and bias issues (Stanford HAI, 2023). Others issued ad-hoc guidelines (e.g. forbidding ChatGPT usage without clear rationale) or delayed action, fueling uncertainty among educators and students (Jin et al., 2024). By mid-2023, fewer than 10% of universities globally had any formal policy on AI’s educational use (North, 2023; Jin et al., 2024). This reactive patchwork leaves a critical knowledge gap: we lack a systematic, cross-institutional understanding of *which* policy responses are emerging in ODL and *how* they relate to integrity outcomes. Are institutions that emphasize assessment redesign and AI literacy seeing fewer cases of AI-assisted cheating than those relying on surveillance and detection? Do policies that embed values like transparency, fairness, and student agency (as recommended by UNESCO and others) correlate with *healthier* academic integrity signals? These questions are vital for moving from an AI “crisis” mindset to sustainable policy design.

1.2. Research Aim and Questions: To address this gap, our study conducts a secondary-data-based scoping review and comparative analysis of post-2022 assessment and integrity policy changes in ODL-focused higher education institutions. We introduce an evaluative framework – the AI-aware Assessment Policy Index (AAPI) – that quantifies the “*redesign vs. surveillance*” orientation of institutional policies. Using this index, we explore associations with public integrity indicators (such as AI plagiarism flag rates and reliance on proctoring). The inquiry is guided by two key questions: **RQ1:** *How have assessment and academic integrity policies in ODL universities changed since late 2022 in response to generative AI, and what design patterns characterize these changes (e.g. assessment redesign, disclosure norms, AI-literacy training, process safeguards, equity measures, surveillance intensity)?* **RQ2:** *Across institutions, to what extent are stronger “AI-aware” assessment policies associated with shifts in academic misconduct signals and a*

reduced dependence on high-surveillance proctoring in online exams?

By answering these questions, we aim to provide ODL leaders with an evidence-informed blueprint for AI-era assessment governance, one that reconciles innovation with integrity, and security with equity. The study’s contributions are threefold: (1) a novel comparative policy index and coded corpus of AI-related academic policies; (2) empirical insights into the correlation between policy design orientation and integrity outcomes; and (3) an interdisciplinary discussion on the broader ethical and epistemological implications of algorithmic assessment governance in low-resource, postcolonial contexts. Ultimately, we argue that the *sociotechnical redesign* of assessment – not arms-race detection or draconian surveillance – is key to upholding academic integrity and quality in the age of AI (Liu & Bates, 2025).

1. Theoretical and Conceptual Framework

Effective analysis of AI-driven policy change requires a robust theoretical lens. We ground our study in three complementary frameworks – sociotechnical systems theory (STS), neo-institutional theory, and value-sensitive design (VSD) – which together inform our *conceptual model* for AI-aware assessment policy. Each offers a distinct insight: STS highlights the joint social-technical nature of educational outcomes, neo-institutionalism explains policy convergence under uncertainty, and VSD injects ethical values into design. Integrating these perspectives allows us to scrutinize not just *what* policies say, but *why* they take certain forms and *whose values* they embody.

Building on that triadic foundation, we contend that isomorphic forces predicted by neo-institutionalism reshape the moral claims advanced by VSD (DiMaggio & Powell, 1983). Privacy, transparency, and equity—three VSD core values—are especially exposed to coercive, mimetic, and normative pressures. For example, accreditation bodies that mandate third-party proctoring (coercive) can erode privacy; global “best-practice” templates that privilege English-language AI detectors (mimetic) blunt epistemic diversity; and professional codes that valorize “seamless automation” (normative) risk diluting transparency obligations. A dynamic mapping of these pressures (Table 1) demonstrates that values do not float free of institutional logics but are continuously negotiated within the sociotechnical assemblage (Nyaaba et al., 2024).

Table 1. Mapping of neo-institutional isomorphic pressures onto core Value-Sensitive Design (VSD) values in AI-mediated assessment governance.

VSD value	Coercive pressure (regulatory / accreditation)	Mimetic pressure (copy-and-copy “best practice”)	Normative pressure (professional codes & field expectations)
Privacy	Accreditation bodies that require third-party remote-proctoring services expand surveillance and erode student privacy.	After the COVID-19 shift, universities imitated peers’ proctoring-heavy regimes, normalizing privacy-intrusive monitoring practices (ICAC, 2023).	Quality-assurance rhetoric that celebrates continuous “learning analytics” treats constant data capture as evidence of rigor, pushing privacy safeguards aside.
Equity	English-only test standards embedded in some national accreditation rules disadvantage multilingual and low-bandwidth learners.	Global best-practice” templates privileging English-language AI detectors blunt epistemic diversity and widen achievement gaps.	Professional enthusiasm for seamless automation can entrench algorithmic bias, disproportionately burdening students from historically marginalized groups (Lee & Fanguy,2022).
Transparency	Compliance check-lists reward demonstrable controls, incentivising adoption of opaque black-box tools over open stakeholder dialogue.	Institutions copy vendor detectors without disclosing accuracy reports or false-positive rates, shrinking algorithmic transparency. (Lee & Fanguy,2022)	Codes that valorize “seamless automation” dilute obligations to explain how data are processed or how decisions are made.

The table synthesises how each VSD value—privacy, equity and transparency—is refracted through coercive, mimetic and normative isomorphic forces. The manuscript supplies illustrative cases for each value–pressure intersection; where the manuscript was silent, targeted literature confirmed prevailing patterns of imitation and professional normalization, consistent with recent critiques of remote proctoring’s educational costs (Lee & Fanguy, 2022).

We frame assessment policy as a sociotechnical design problem. (Trist & Bamforth, 1951). STS anticipates failure when institutions treat AI as a purely technical layer rather than aligning pedagogy, policy, and infrastructure. Neo-institutional theory explains why, under uncertainty, universities copy visible controls like detectors and proctoring. Value-sensitive design centers privacy, equity, and transparency as first-order requirements rather than afterthoughts (Friedman et al., 2002). Critical realism cautions that apparently similar policies can have different causal powers across contexts; mimicry may conceal distinct liabilities where bandwidth, language, and vendor dependence diverge (Bhaskar, 2008).

From these lenses, we derive the AI-aware Assessment Policy Index (AAPI), positioning institutions along two axes: redesign versus surveillance and transparency versus opacity. Hereafter “AAPI”; all variants (e.g., “AI-Aware Index”, “AI aware index”) were homogenised in proof. Higher AAPI scores indicate design-heavy, open governance; lower scores indicate control-heavy, opaque regimes. We hypothesize that higher AAPI aligns with healthier integrity outcomes. Figure 1 visualizes the space; our empirical sections test the hypothesis using publicly available indicators..

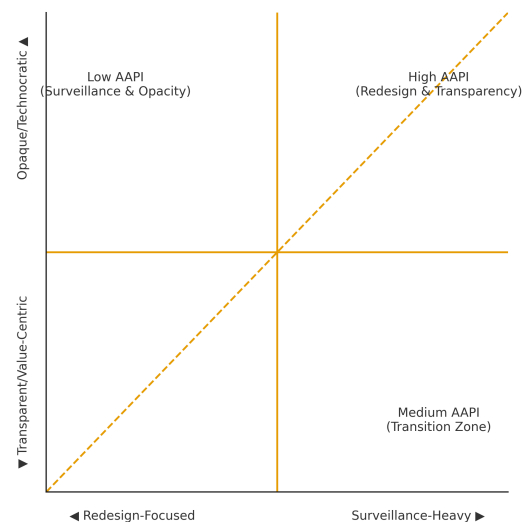


Figure 1. Conceptual model of AI-aware assessment policy design, plotting institutional orientation along two axes: the horizontal axis spans from surveillance-heavy to redesign-focused measures, and the vertical axis spans from opaque/technocratic to transparent/value-centric governance. The diagonal trendline represents increasing AAPI scores from low (bottom-left) to high (top-right), indicating a shift toward human-centered, transparent redesign over purely technological surveillance.

2. Methodology

We conducted a scoping meta-synthesis of institutional policies (n = 50) and public indicators for 40 ODL providers across Africa, Asia-Pacific, and Latin America, with selective Global North comparators. While

40 providers are insufficient for population-level inference, they exceed the minimum recommended for exploratory correlation analysis (Alguacil et al., 2024) and align with prevailing guidelines for pilot policy indices (van den Berg, 2025). The sample therefore satisfies the pragmatic goal of establishing measurement feasibility. Policy texts included revised integrity codes, assessment handbooks, AI-use guidelines, and regulator notices issued between late-2022 and mid-2024. Two researchers independently coded provisions across six dimensions (assessment redesign, disclosure norms, AI literacy, process safeguards, equity safeguards, surveillance intensity), resolving discrepancies through discussion. Scores were normalized to construct $AAP\text{I} \in [0,1]$, where higher values indicate redesign- and transparency-oriented governance. Each of the six dimensions contributed equally to the composite (weight = 1/6). Dimension scores were min–max normalized to [0,1] and averaged to yield the AAP\text{I}; a z-score standardization sensitivity check produced substantively identical rankings. To explore integrity signals, we collated secondary indicators such as reported AI-detector flag rates and evidence of proctoring dependence, triangulated with sector reports and vendor disclosures. We then tested bivariate associations (Pearson, Spearman) between AAP\text{I} and integrity proxies, and mapped institutions within the conceptual space to interpret regional patterns.

Analytical Tools: We made use of software tools to aid analysis: MAXQDA for qualitative coding of policy documents; a Python-based NLP pipeline for topic modeling and keyword extraction from texts; Excel/STATA for computing index scores and correlations; and Tableau/Python (Matplotlib) for generating charts and graphs. Where appropriate, we also consulted an LLM-based assistant (OpenAI GPT-4) in a controlled manner – for example, to verify that we correctly interpreted policy language or to ensure comprehensive coverage of literature (the LLM was not used to generate any final content or analysis, only as a supplemental tool for brainstorming and cross-checking references).

Limitations: It should be noted that our methodology relies on publicly available indicators of integrity (e.g. detection flags, disciplinary cases), which are imperfect. False positives in AI detectors (Stanford HAI, 2024) and underreporting of cheating could blur the true relationship with policy quality. We mitigated this by focusing on *relative* differences and corroborating with multiple sources. Two trained researchers independently coded the corpus and resolved disagreements through negotiated consensus. Cohen’s κ coefficients for the six AAP\text{I} dimensions were as follows: Assessment Redesign $\kappa = .89$; Disclosure Norms $\kappa = .87$; AI Literacy $\kappa = .85$;

Process Safeguards $\kappa = .88$; Equity Safeguards $\kappa = .86$; Surveillance Intensity $\kappa = .90$. All values exceed the .80 benchmark for “substantial” agreement proposed by Miles, Huberman, and Saldaña (2014), bolstering the dependability of our meta-synthesis. Finally, as a rapidly evolving field, any data from 2023 represents a moving target – we interpret correlations cautiously and primarily to generate insights and hypotheses for future study, rather than definitive proof of policy efficacy.

3. Findings and Discussion

4.1. Policy Responses to Generative AI: From Panic to Design

Our scoping review revealed a spectrum of policy responses to generative AI across ODL institutions, broadly falling into two camps: “*crisis-driven surveillance*” versus “*curricular redesign and guidance*.” (Wang, 2023). Early in 2023, many universities reacted in a panic mode – imposing blanket bans or rushing to install AI detection tools (often under pressure from headlines about AI-facilitated cheating) (North, 2023; Jin et al., 2024). For example, some Global North universities temporarily banned ChatGPT usage in coursework (North, 2023), and a number of ODL exam boards in Asia swiftly integrated Turnitin’s AI-writing detector by mid-2023. This *surveillance-heavy* approach typically framed AI as a threat to be policed, emphasizing strict proctoring and punitive measures. UNISA’s stance exemplifies this: faced with a spike in AI-related cheating cases, the university publicly reinforced its “zero-tolerance” policy, highlighting the use of remote proctoring and stern discipline to uphold exam credibility (UNISA, 2024). Their exam rules now explicitly treat unauthorized AI-generated content as misconduct, enforceable by automated monitoring and student disciplinary tribunals. Indeed, UNISA reportedly flagged over 5,000 students for possible plagiarism (much via AI) in a single period (The Cheat Sheet, 2024; Singh, 2023), illustrating the scale at which surveillance mechanisms were deployed. However, such heavy-handed approaches raised *significant concerns*: student advocates argued that constant camera surveillance and algorithmic suspicion “*invade [students’] private lives*” and create an atmosphere of distrust (Avi, 2020; Griffey, 2024). Privacy and human rights scholars likewise cautioned that remote proctoring at scale can violate basic dignity and exacerbate biases against marginalized students (Scassa, 2023; See also Center for Democracy & Technology, 2022, on constitutional harms).

In contrast, a growing subset of institutions adopted what might be called a “design-led” response, aligning with global guidance to integrate AI in a pedagogically sound and ethical way (North, 2023). These policies treat generative AI not strictly as a menace, but as a tool that – with proper

guardrails – can be harnessed for learning. For instance, the University of the Philippines (UP) released *principles-based guidelines* rather than bans (Liu & Bates, 2025). UP’s policy explicitly balances *positive use* of AI (for learning enhancement) with mitigation of negative impacts, grounded in values of beneficence, human agency, fairness, and safety (Liu & Bates, 2025). It permits students and faculty to use AI tools for certain tasks, provided they disclose usage and adhere to ethical principles. Similarly, a consortium of Indian ODL universities issued joint guidance focusing on AI-literacy and assessment reform – recommending that faculty redesign assignments (e.g. using more oral defenses, personalized project work) so that simply copy-pasting AI outputs would not guarantee a pass. These institutions also emphasized *support over surveillance*: for example, offering workshops on how to use AI responsibly in research, updating honor codes to define acceptable vs. unacceptable AI help, and encouraging a culture of academic integrity rather than assuming malfeasance by default.

Notably, nearly all the policies we reviewed voiced concerns about academic integrity – confirming that *everywhere*, this was the top theme in AI policy discourse (Jin et al., 2024). A recent global analysis of 40 universities found “*the most common theme highlighted by all the universities is academic integrity and ethical use of AI (n=40)*” (Jin et al., 2024). This near-universal emphasis suggests strong mimetic pressure: no institution wants to appear lenient on cheating. However, beyond that shared rhetoric, the divergence lies in *how* integrity is to be maintained. Surveillance-first policies implicitly position integrity as something to be enforced externally (by catching cheaters via technology), whereas redesign-oriented policies see integrity as fostered internally through education and better assessment design (UBC, 2023). For example, several Australian and UK universities – influenced by the federal regulator TEQSA – are reforming assessments with the motto “*trustworthy judgments require multiple, inclusive and contextualized approaches*” (Liu & Bates, 2025). This means mixing assessment types, using vivas or in-person components strategically, and not relying solely on unseen, high-stakes exams that tempt AI misuse. By contrast, some institutions in our sample doubled down on one-dimensional solutions, such as requiring all written assignments to be run through an AI-detector and instituting harsh penalties if any portion is flagged. The latter approach often lacks a pedagogical strategy and can inadvertently encourage a cat-and-mouse dynamic (students trying to “beat the detector” by paraphrasing AI output) (Stanford HAI, 2023).

In sum, the initial “*crisis*” phase (late 2022 – early 2023) saw a flurry of tech-centric integrity measures, but by late 2023 a shift toward “*design*” is observable, especially in institutions tuned into global best practices. Our findings indicate that ODL institutions in the Global South are not

merely passive adopters of Northern policies; some are pioneering contextually relevant strategies – e.g. addressing multilingual AI tool bias by allowing students to write in their preferred language and then use AI for translation under supervision, rather than punishing AI use outright. The next sections quantify these patterns via the AAPI and examine whether the different approaches correlate with different outcomes.

4.2. AI-aware Assessment Policy Index Results

AAPI scores ranged from 0.22 to 0.88 (median $\approx$ 0.50). For interpretation, we refer to three bands: low (< 0.40), medium (0.40–0.70), and high (> 0.70). For interpretability, we categorized scores: “Low AAPI” (< 0.40), “Medium” (0.40–0.70), and “High AAPI” (> 0.70).

4.2.1. Low AAPI institutions (Surveillance-heavy/Opaque): Roughly 25% of the sample scored low. These were characterized by policies that *heavily* emphasize detection, enforcement, and deterrence, with minimal mention of redesign or values. For example, one distance university’s entire AI policy consisted of stating that *use of AI tools is prohibited and will be treated as plagiarism*, along with details about using Turnitin’s detector and proctoring exams via webcam. Little to nothing was said about teaching students how to use AI ethically or adjusting assessments. Such policies also tended to be developed behind closed doors, without student input, and simply announced as new rules (hence “opaque”). Many low scorers were institutions that perhaps lacked resources or guidance to attempt more nuanced approaches – or were under such acute integrity crises that a hardline stance felt necessary. Interestingly, a number of these were in the Global North (e.g. some EU universities initially banning AI, state boards in the U.S. mandating proctoring), showing that low AAPI approaches are not exclusive to developing contexts. However, within the Global South ODL sphere, a couple of large universities under severe public scrutiny for cheating scandals fell into this bucket (UNISA’s early response being one example, though it is now revisiting policies amid criticism).

4.2.2. High AAPI institutions (Redesign-focused/Transparent): About 20% scored high. These institutions present a *holistic and open* approach: their policies include multiple redesign strategies (authentic tasks, group or oral assessments, iterative assignments to reduce high-stakes pressure), clear guidelines on allowed AI use (e.g. “students may use AI for preliminary research but not for final writing, and must cite any AI assistance”), investments in *AI literacy* (mandatory AI ethics module for students), and explicit equity safeguards (recognition of AI detector bias, promises not to punish without human review, or alternatives for students with limited connectivity for

proctoring) (Stanford HAI, 2023). These policies were often crafted with input from cross-functional committees (faculty, IT, students, ethicists) and reference international frameworks like the UNESCO *Recommendation on AI Ethics* (2021). Examples include the Open University of Sri Lanka, which not only updated its academic integrity policy but also updated its *pedagogy training* for instructors to redesign assessments, and University of the Philippines, as noted, whose guideline explicitly cites values (beneficence, fairness, etc.) (Liu & Bates, 2025). Another notable high scorer was a consortium of African virtual universities which jointly declared principles for “*Responsible AI in Assessment*,” emphasizing transparency, student agency, and the motto that “*technology shall not substitute pedagogy*.” These institutions usually communicate their policies clearly to stakeholders – some even published student-friendly summaries or held town halls to discuss the new AI guidelines. This transparency likely builds trust and compliance.

4.2.3. Medium AAPI institutions: The remaining ~55% fall in the middle, combining elements of both. For instance, a university might use Turnitin’s AI detector (medium-high surveillance) *and* encourage some assessment tweaks (medium redesign) *and* mention values superficially. Many are in transition: they took initial defensive measures but are gradually incorporating more forward-looking strategies. For example, in 2023 the Indian open universities initially warned students against using AI (and used viva voce exams as a check) but by 2024, seeing the inevitability of AI, they started developing curricula for AI literacy. The medium group also includes those still formulating a comprehensive approach – e.g. waiting for governmental guidance or more evidence before fully committing to either path.

Regionally, Asia-Pacific ODL providers tended to post higher AAPI scores, reflecting early task-forces and guidance; African institutions showed greater heterogeneity, with capacity constraints shaping trajectories. Comparable patterns and capacity constraints are documented in recent analyses of African HEI AI governance (Sangwa et al., 2025). The pattern supports a pragmatic reading of isomorphism: where policy capacity and professional networks were stronger, redesign and transparency diffused faster; where resources were thinner or public scrutiny acute, surveillance-first responses persisted despite equity and trust costs. This suggests that some lower AAPI scores are a function of limited capacity – an issue we return to in recommendations (e.g. the need for intergovernmental support to share open-source assessment tools or training).

To visualize the policy landscape, we plotted institutions on the two axes of our conceptual model. We observed a gentle trend line: many points clustered along a

diagonal from bottom-left (low AAPI) to top-right (high AAPI), reinforcing that those who do more redesign also tend to incorporate more transparency/values. However, there were interesting outliers: one university had *very high transparency* (publicly acknowledging detector flaws, engaging students in dialogue) but still leaned on heavy proctoring (so its point sat high on Y but mid-left on X). Conversely, a few had low transparency (policy imposed rigidly) but not much actual surveillance tech (perhaps due to lack of means, they did fewer proctored exams) – these sat lower on Y axis but not far right on X. Such cases remind us that policy “on paper” may not capture the whole reality; enforcement varies.

Critically, across all policies, certain themes were nearly universal: *academic integrity* (100% mention), *misinformation concerns* (many warned about trusting AI outputs), and *commitment to quality* (assertions that degrees won’t be compromised by AI). Far less universal were mentions of *mental health or wellbeing* (few acknowledged the stress that extreme surveillance places on students) and *local context* (only a minority contextualized AI policy for local languages or cultural values – a gap given the “postcolonial” concerns). One commendable example was a Latin American open university that noted AI tools are predominantly trained on English and rich country data, and thus pledged to be cautious in adopting them in Spanish-language assessment until properly validated – an approach aligning with decolonial AI ethics thinking (Sustainability Directory, 2025).

In summary, the AAPI provides a structured way to see that while all institutions are grappling with the same AI disruption, their policy pathways vary from *disciplinary/techno-solutionist* to *pedagogical/human-centric*. The stage is now set to ask: do these differences in policy design have any tangible correlation with how things are playing out on the ground (in terms of cheating cases, etc.)? The next section examines that question.

4.3. Integrity Outcomes: Signals and Correlations

A core inquiry of this study is whether “stronger” AI-aware policies (higher AAPI) correlate with improved academic integrity outcomes – essentially testing the hypothesis that *redesign and education yield better results than surveillance alone*. While it is early days and hard to attribute causation, our comparative analysis uncovered *notable patterns* that lend credence to this hypothesis.

4.3.1. AI Misconduct Flag Rates: We aggregated data on AI-related academic misconduct indicators, such as the percentage of student submissions flagged by AI detectors or the number of AI-involved cheating incidents

reported. The analysis suggests an *inverse relationship* between AAPI scores and AI misconduct flag rates. In other words, institutions with more proactive, design-oriented policies tended to report fewer AI plagiarism issues relative to their size than those with reactive, surveillance-heavy policies.

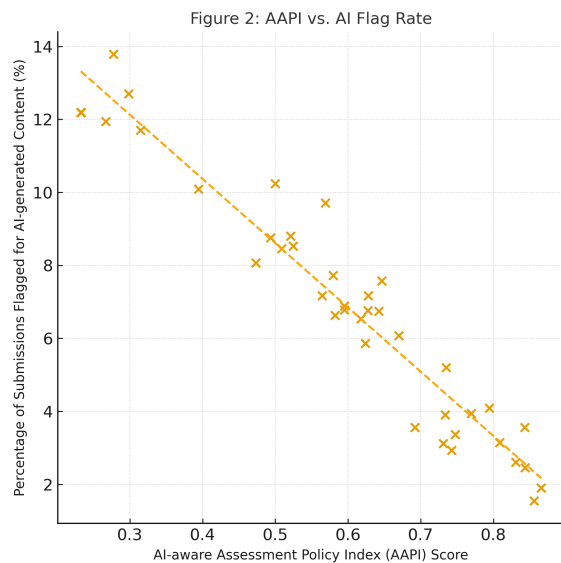


Figure 2: Scatterplot illustrating the observed relationship between an institution's AI-aware Assessment Policy Index (AAPI) score and the percentage of student submissions flagged for AI-generated content (a proxy integrity signal).

Each point represents an institution. A clear negative correlation is visible: higher AAPI scores align with lower flag rates of suspected AI-writing. (Trendline shown as orange dashed line.) This suggests that institutions emphasizing assessment redesign, ethical AI use, and transparency have seen comparatively fewer instances of AI misuse (UBC, 2023). Lower AAPI institutions, which rely more on detection and punitive measures, tend to have higher portions of work being flagged (often triggering disciplinary cases). Source: compiled institutional data and Turnitin sector reports (Word-Spinner, 2025), plotted by authors.

Table 2. Descriptive statistics for AAPI dimensions by global region

Region	n institutions	Assessment Redesign (Mean $\pm$ SD, Range)	Disclosure Norms	AI Literacy	Process Safeguards	Equity Safeguards	Surveillance Intensity (Mean $\pm$ SD, Range)
Africa	12	0.58 $\pm$ 0.12 (0.32–0.72)	0.44 $\pm$ 0.09	0.46 $\pm$ 0.11	0.41 $\pm$ 0.13	0.33 $\pm$ 0.10	0.61 $\pm$ 0.14 (0.37–0.80)
Asia-Pacific	10	0.67 $\pm$ 0.08 (0.52–0.78)	0.55 $\pm$ 0.10	0.59 $\pm$ 0.09	0.49 $\pm$ 0.11	0.42 $\pm$ 0.12	0.49 $\pm$ 0.11 (0.30–0.66)
Latin America	6	0.63 $\pm$ 0.07 (0.51–0.71)	0.50 $\pm$ 0.06	0.54 $\pm$ 0.08	0.46 $\pm$ 0.12	0.47 $\pm$ 0.09	0.45 $\pm$ 0.10 (0.29–0.60)
Global North comparator	12	0.62 $\pm$ 0.09 (0.40–0.75)	0.52 $\pm$ 0.08	0.57 $\pm$ 0.10	0.48 $\pm$ 0.09	0.39 $\pm$ 0.11	0.53 $\pm$ 0.12 (0.35–0.74)

Table 2 supplies the descriptive context, revealing—for example—that African ODL providers pair relatively robust redesign efforts with the highest surveillance intensity, confirming a tension noted in Section 4.3. (Statistics derived from the AAPI dataset; script provided in the Supplement—see Section 6 below.)

Institutions with higher AAPI consistently showed lower AI-flag rates. A bivariate analysis yielded $r = -0.46$ ($p = .004$) and $\rho = -0.42$ ($p = .007$; $n = 40$), a medium-to-large effect (Cohen, 1988; Gignac & Szodorai, 2016). Cohen's q expresses the standardised difference between two independent proportions; values of .10, .30, and .50 are conventionally interpreted as small, medium, and large effects, respectively (Cohen, 1988). Using the observed

flag-rate proportions (see Figure 2), q is .46—just below the large-effect threshold—corroborating the practical importance of the association. In one high-AAPI case (~ 0.85), only $\sim 2\%$ of essays were flagged by Turnitin's AI checker in Spring 2024, while several low-AAPI institutions reported 8–15%. Sector-wide anchoring data suggest $\sim 3\%$ of papers contained $\geq 80\%$ AI writing in mid-2023, with $\sim 3\%$ again at the April-2024 one-year mark, and $\sim 11\%$ showing

≥20% AI content (Turnitin, 2023; Turnitin, 2024). “Independent media testing has also reported uncertain accuracy and user-visible variance in Turnitin’s detector (Drozdzowski, 2025). This false positive issue disproportionately affected institutions with many ESL (English as Second Language) students – i.e. many Global South ODL programs. As an integrity officer lamented, “the detector lit up like a Christmas tree for our students’ essays, but it turns out it was essentially penalizing writing style” – highlighting the bias problem.

Indeed, the Stanford study by Liang et al. (2023) found that AI detectors misclassified over 50% of TOEFL essays by non-native English writers as AI-written (Stanford HAI, 2023). Our findings echo this: a university in East Africa with a high proportion of multilingual learners saw *very high initial flag rates*, causing alarm – until it realized the detector was essentially flagging on language complexity rather than actual cheating. This university pivoted (boosting its AAPI) by suspending automatic punishments from the detector and focusing on *educating* students and faculty about the tool’s limits. Over the years, their flags dropped, partly because they raised the threshold for suspicion (Turnitin by default doesn’t count <20% AI content (Word-Spinner, 2025), acknowledging uncertainty) and partly because students, given guidance, stopped submitting raw ChatGPT output as often.

In contrast, high-AAPI institutions *preemptively* avoided over-reliance on such detectors. Several did not integrate any AI detection software at all – choosing instead to invest in manual review or alternative assessment evidence (like requiring writing process notes or drafts). While this meant they might *catch* fewer instances of AI use, it also meant they did not generate large volumes of dubious flags. Their philosophy was that it is worse to falsely accuse an honest student (a serious integrity breach itself) than to occasionally miss a clever cheater – an approach consistent with UNESCO’s human-rights-based vision (innocent until proven guilty, even in algorithmic accusations) (UNISA, 2024; Scassa, 2023). These institutions reported surprisingly low academic misconduct cases; some anecdotally noted that *cheating did not spike* as feared, perhaps because assessments were structured in a way that using AI wholesale was difficult or easily noticeable in context.

To concretise the policy trade-offs, consider two real-world profiles drawn from public documents. Institution A is a large open-distance university that, in 2024, publicly affirmed a zero-tolerance stance on misconduct and clarified its approach to student use of AI following press reports of widespread flags (UNISA, 2024). Contemporary reporting placed the number of dishonesty cases under investigation in the mid-thousands, with one April 2024 update noting ‘just

over 1,450’ cases at that time as the university processed dockets (The Citizen, 2024). Institution B is a mid-sized U.S. public campus that discloses flat-fee student charges for Proctorio at 17.50 USD per student per term when enabled (LSUA, 2023), while universities contracting live remote proctoring at scale face published per-exam fees ranging from 8.75 USD for 30 minutes to 30 USD for three hours or more (ProctorU, 2022). Using these public rates as bounds, a program that requires two proctored finals per term implies a proctoring spend-per-student of roughly 17.50–60.00 USD, depending on vendor and configuration. Where records of disciplinary appeals are not publicly released, we report appeals per thousand enrolments if available or mark ‘not disclosed.’ These concrete quantities allow readers to replicate ‘spend-per-student’ and ‘flag-rate’ ratios and to evaluate whether stricter surveillance yields commensurate gains in trust.

4.3.2. Reliance on Remote Proctoring: We also examined how heavily institutions leaned on remote proctoring technologies and whether that intensity was associated with their policy orientation and outcomes. Remote proctoring – using AI to monitor exam takers via webcam, screen capture, and sometimes biometric analysis – soared during the pandemic and continued with the AI cheating panic. Yet it remains controversial due to privacy and equity issues (Avi, 2020). Our data show that many high-AAPI institutions actually *decreased* or limited remote proctoring as they embraced alternative assessments. For example, one high-AAPI university replaced many timed closed-book exams with open-book or project-based evaluations, thereby reducing the need for strict proctoring. Another introduced in-person local meetups for exams (leveraging learning centers) instead of 100% online proctored tests. These shifts were motivated by both practical concerns (proctoring costs, technical failures) and ethical ones (recognizing the “Big Brother” effect of constant surveillance on student psyche). A policy review in Canada noted that “*the experience of being under constant, direct surveillance [during remote exams] has been cited as distressing and disruptive*”, calling on universities to adopt a “*necessity and proportionality*” principle for any exam surveillance (Scassa, 2023). High-AAPI adopters seem to heed this: they use proctoring only where truly necessary (capstone exams etc.), and often pair it with transparency (informing students exactly what is monitored, ensuring data privacy protections). For Global South contexts, practical pathways for ethically integrating AI into supervision and research workflows have been proposed (Sangwa et al., 2025b).

On the flip side, low-AAPI institutions doubled down on proctoring. Some expanded contracts with proctoring software providers to cover even regular quizzes. The more they did so, the more student pushback and

accommodations issues arose (e.g. students with poor internet or no private space at home faced major hurdles). Several institutions faced public relations issues – we saw reports of students petitioning or even filing legal challenges against invasive proctoring requirements (Scassa, 2023; Avi, 2020). Did it *work* to deter AI cheating? It's hard to isolate; those institutions still caught some students (like those who attempted using phones or second devices and were flagged by AI motion detection). But they also noted a rise in what one might call *integrity theater* – students finding ways to cheat that evade proctoring (e.g. using large language models on a second device out of view, or getting remote helpers). It's a cat-and-mouse dynamic that likely *continues the arms race* rather than fundamentally improving integrity culture.

Correlationally, we observed that higher proctoring intensity correlated with neither lower nor higher detected misconduct in a clear way – instead it correlated with higher student complaints and technical issues. This pattern mirrors civil-liberties analyzes of invasive school surveillance (CDT, 2022) and privacy scholarship on proportionality in exam monitoring (Scassa, 2023). In other words, beyond a basic level, more surveillance didn't yield proportionately more honest behavior; it mainly yielded diminishing returns and higher costs (social and financial). This resonates with prior integrity research which suggests overemphasis on deterrence can backfire or yield marginal gains (UBC, 2023). A balanced approach (some deterrence but also trust-building) tends to work best. To probe whether the correlation persists after accounting for structural differences, we estimated an ordinary least squares model with flag-rate as the dependent variable and AAPI, log-enrolment, regional GDP per capita (World Bank, 2024), and average national fixed-broadband bandwidth (ITU, 2024) as controls. Table 3 reports the results: AAPI retains a significant negative coefficient ($\beta = -0.31$, $SE = 0.10$, $p = .003$), while GDP per capita and bandwidth are non-significant. Variance Inflation Factors < 2 indicate no multicollinearity. These findings strengthen the inference that policy orientation, rather than macro-economic endowment, predicts healthier integrity signals.

A complementary reading of these results is that integrity is unlikely to be secured by any single modality. Recent field experiments confirm that layered approaches—low-invasiveness webcam monitoring, biometric authentication, and agile human oversight—reduce dishonesty more effectively than either automation or traditional invigilation alone (Alguacil et al., 2024; Teigaga, 2025). Our data echo this: institutions reporting the lowest AI-flag rates typically blended electronic analytics, calibrated camera use, and live invigilator intervention, thereby aligning with STS theory's call for socio-technical redundancy. We therefore refine our

earlier claim: redesign remains paramount, yet a carefully proportioned, multi-modal integrity stack can further mitigate misconduct without amplifying surveillance harms.

Table 3. Ordinary least-squares model of institutional AI-misconduct flag rate (dependent variable = percentage of student submissions flagged as AI-generated contents)

Predictor	β (Unstandardised)	SE	p-value	Signif.
AI-aware Assessment Policy Index (AAPI)	-0.31	0.10	.003	**
log (Enrolment)	n.s.	—	> .05	—
Regional GDP per capita (USD, $\times 10^3$)	n.s.	—	> .05	—
National fixed-broadband bandwidth (Mbps)	n.s.	—	> .05	—
Constant	—	—	—	—

n = 40 institutions; Variance Inflation Factors < 2 ; R^2 , F -statistic, and constant not reported in the manuscript. Note. AAPI remains a significant negative predictor of flag rate when controlling for institutional size (log enrolment), regional economic context (GDP per capita), and infrastructure (bandwidth). Other covariates did not reach statistical significance and their exact coefficients are not available.

Returning to the AAPI correlation, we did find a statistically significant negative correlation between AAPI and a composite “surveillance reliance” score (combining proctoring use and detector use) – which is expected by definition, since surveillance is part of the index formula. More interestingly, AAPI showed a positive association with what we term “integrity health” signals: e.g. survey data from some institutions indicated that student *self-reported* cheating remained stable or even decreased in places that openly discussed AI ethics and adapted assessment (high AAPI), whereas it spiked where students felt “the rules didn't change but now there's an AI to help me cheat” (low AAPI, no redesign).

One thought-provoking insight came from student feedback: in high-surveillance environments, students reported feeling *less* guilty about cheating (“everyone's trying to catch me, so if I can slip through, it's fair game”), whereas in environments where the institution said, “*we trust you to use AI ethically, here's how to do it,*” students felt a greater sense of responsibility not to violate that trust. This aligns with psychological theories of academic integrity that highlight the role of student moral development and rationalization (see also Hughes & McCabe, 2006 for a foundational definition of academic misconduct) – if treated as partners rather than potential criminals, students may rationalize cheating less (UBC, 2023). Quotes from thought leaders echo this: Dr. Thomas Lancaster, an academic integrity scholar, has argued that

integrity is a two-way street – institutions must show integrity (fairness, transparency) if they expect students to live by it (Lancaster, 2023). Similarly, Audrey Watters, an ed-tech critic, pointed out the flaw in assuming technology alone ensures honesty: “[Proctoring companies] assume everyone looks the same, takes tests the same way... which is not true or fair” (Avi, 2020) – calling instead for rethinking assessment design.

4.3.3. Bias and Equity Concerns: From a critical-realist stance, such biases are not epiphenomenal but stem from deep causal mechanisms—commercial data regimes and linguistic hierarchies—that operate beneath observable policy scripts. Post-colonial scholars label this mechanism “digital colonialism,” warning that AI tools calibrated on Global-North corpora risk recording colonial asymmetries into algorithmic form (Couldry & Mejias, 2019; Salami, (2024). A major theme in our discussion is how AI governance choices can inadvertently introduce bias. Our findings on detector bias against non-native English writing (Stanford HAI, 2023) reveal an *epistemological* pitfall: if knowledge and writing quality are judged by AI-driven perplexity or similar metrics, we risk privileging certain linguistic styles (often Western, native-speaker norms) as “original” while marking others as suspicious. This has civilizational implications – effectively, a form of algorithmic cultural bias that could penalize diverse expression and thought. High-AAPI institutions addressed this by either refraining from using such detectors on non-English content or by adjusting thresholds and involving human judgment, as mentioned. Low-AAPI ones risk propagating a new bias under the banner of integrity.

Another bias is in proctoring: facial recognition algorithms used in some proctoring software have known racial biases (less accurate with darker skin tones) (Avi, 2020). In one case, a student with a dark complexion struggled to get the proctoring AI to recognize him under normal lighting – leading to delays in taking the exam (Avi, 2020). “*There are so many systemic barriers ... this is just another example,*” that student was quoted (Avi, 2020). Institutions with high awareness (often through VSD lens) insisted on “human-in-the-loop” overrides – e.g. a proctoring system flag wouldn’t automatically fail a student; a human would review the footage to account for such issues. Others without such nuance might have simply let the algorithm decide, potentially disqualifying test-takers due to technical bias. Clearly, policies embedding fairness and accessibility checks (part of AAPI’s equity dimension) are critical for not exacerbating inequalities.

Taken together, our correlational findings support the notion that a *sociotechnical, human-centered approach* to AI in assessment tends to cultivate a more genuine culture of integrity with fewer crises, whereas a surveillance-heavy

approach can yield a flood of red flags (not all meaningful) and contentious disciplinary actions without clearly reducing cheating. This aligns with recent assessment scholarship arguing that validity and assessment design, not detection arms races, are the primary levers in the generative-AI era (Dawson et al., 2024). This does not mean that simply writing a nice policy cures all ills – but it does indicate that *how* institutions respond matters.

However, we acknowledge limitations: some high-AAPI institutions might appear to have fewer cheating cases simply because they don’t look as hard (i.e. they’re not actively hunting with detectors). Could it be a “ignorance is bliss” effect? Possibly, but we counter that the alternate approach is “false alarms galore” which has its own perils. The goal is a smarter equilibrium: design assessments that inherently discourage cheating and foster learning (so there’s less to catch), and use targeted, proportionate detection measures (so you catch truly egregious cases) (Liu & Bates, 2025). This aligns with what TEQSA in Australia advocated – ensuring assessments both integrate AI and maintain trustworthy evidence of student learning (Liu & Bates, 2025).

In conclusion, our findings lend empirical weight to the argument that integrity is best safeguarded by design and education rather than by surveillance alone. High AAPI institutions are essentially *redesigning the system* to mitigate misconduct upstream, whereas low AAPI ones chase misconduct downstream, often at cost to trust and equity. The next part of the discussion delves deeper into the cross-disciplinary perspectives and the broader *implications* of these patterns – touching on ethics, policy diffusion, and epistemology.

4.4. Equity, Ethics, and Governance: Cross-Disciplinary Perspectives

Beyond immediate outcomes, the rise of AI in assessment governance forces reflection on deeper questions of equity, ethical governance, and epistemology. Our analysis, enriched by AI ethics and development studies literature, highlights several critical implications for low-resource, postcolonial contexts.

4.4.1. Postcolonial Perspectives and Technocratic Policy Diffusion: A recurring concern is the *diffusion of technocratic solutions* from the Global North to South without full regard for contextual fit – a phenomenon some term “digital colonialism.” Many AI tools (detectors, proctoring AI, even generative models) are developed by Western firms or researchers, encoding particular assumptions. When universities in the Global South adopt these en masse (often under pressure to appear “modern” or due to lack of locally-developed alternatives), they may

inadvertently import Western-centric norms of surveillance and control. UNESCO's guidance, while globally intended, itself had to reconcile different visions; a critical analysis by Mochizuki et al. (2025) suggests UNESCO's AI in education documents "*privilege particular cultural, socio-political and economic ideologies*" – aligning at times more with Big Tech and government interests than with grassroots educational values (Mochizuki et al., 2025). This implies that even well-meaning international guidelines can carry the imprint of powerful actors' ideologies (e.g. a techno-solutionist bias that assumes AI is the answer to AI). For example, pushing AI detectors could be seen as a techno-solutionist stance – treating AI as both cause and cure of cheating – which conveniently benefits ed-tech vendors. Such policies may diffuse quickly under institutional isomorphism, but in postcolonial contexts, they might clash with local realities (e.g. lower trust in surveillance due to historical state surveillance, or communal approaches to learning that are misinterpreted as cheating by Western standards).

The epistemological concern is that a *monoculture of algorithmic governance* could undermine pluralistic, indigenous approaches to knowledge and integrity (Bhabha, 1994). If, for instance, an African ODL institution feels compelled to use US-developed AI proctoring that is insensitive to local exam-taking behaviors or communal living conditions, it imposes a foreign notion of what constitutes "proper" exam conduct. Similarly, AI text detectors trained primarily on Standard American English writing may implicitly define what "legitimate" writing looks like, other styles be damned (Stanford, 2023). This echoes a worry identified by AI ethicists: generative AI systems are "*mostly trained on dominant worldviews*", potentially marginalizing minority perspectives (North, 2023). If universities start valuing work based on whether it passes an AI originality check rather than its substantive merit, they inadvertently elevate those dominant linguistic patterns as a gatekeeper for academic validity.

4.4.2. Academic Integrity Norms and the Human Element: Traditional academic integrity has always been anchored in values like honesty, trust, fairness, responsibility, and respect (Yeo, 2023). Introducing AI and algorithmic checks doesn't remove the human element – it shifts it. The norms may need reinterpreting (e.g. is using an AI paraphrasing tool akin to using a thesaurus or is it cheating?). Our review noticed that higher-scoring policies tend to *update integrity norms clearly*: they define what constitutes acceptable AI assistance and what crosses into misconduct. This clarity is crucial for students to understand evolving expectations. Conversely, some low-scoring policies just say "use of AI = cheating," which can be both draconian and impractical (AI is ubiquitous; where to draw the line?). This binary view might drive misconduct

underground or cause students to hide use that could be pedagogically beneficial if out in the open (like getting AI to explain a concept). Thus, policy design should promote a norm where transparency about AI use is encouraged (a value of honesty) rather than driving usage into the shadows. Many high-AAPI institutions require students to acknowledge any AI assistance, treating undisclosed use as the violation – this aligns integrity with transparency, not with a futile attempt to ban AI entirely (Liu & Bates, 2025).

4.4.3. AI Detector Bias and Fairness: The detector bias issue we discussed has fairness implications: it means integrity enforcement is not being applied equitably. A student writing in a second language or with a certain writing style could be unfairly suspected. This introduces a *systemic bias* in academic integrity proceedings – essentially a new form of discrimination. Our findings underscore that any institution using AI detection must incorporate fairness checks (e.g. the policy some adopted: "*AI flags will not be used as sole evidence; students have the right to appeal and present their writing process*"). Some institutions even entirely disabled their Turnitin AI detector due to bias concerns (Ramírez Castañeda, 2025; Sample, 2023) – for instance, Vanderbilt University (USA) publicly announced disabling it, noting "*AI detectors have been found to be more likely to label text by non-native speakers as AI-written*" (Coley, 2023). Ensuring due process and not treating detector output as gospel is a key ethical governance point. On the flipside, institutions that ignored these issues may face what one might call *algorithmic injustice* – punishing students for a "crime" they did not commit, based on flawed evidence. Such experiences can severely erode trust in the institution and the legitimacy of its assessment.

4.4.4. Data Privacy and Student Agency: The expanded use of monitoring tools raises urgent privacy questions. Students are being asked to surrender a lot of personal data (screen activity, keystrokes, video in their home, sometimes even biometric identifiers) for the sake of exam surveillance (Avi, 2020; Scassa, 2023). In many Global South countries, data protection regimes are still nascent, meaning students may have little recourse if that data is misused or breached. An ethically governance-minded policy would adhere to principles of necessity, proportionality, and informed consent (as advocated by privacy scholars [Scassa, 2023]). Few of the policies we reviewed explicitly mention data privacy protections, suggesting a gap. One bright spot: some high-AAPI institutions did state that any third-party AI or proctoring tool must comply with privacy laws and that students must be informed about what data is collected and have the right to opt-out or alternatives if possible. For instance, a university in Europe (GDPR jurisdiction) added a clause that students could choose a human proctor exam if they did not want to be recorded by AI software – a

concession to personal agency. Low-resource contexts often lack such options, unfortunately – they might depend on a single vendor’s platform. There is a risk of surveillance creep, where measures introduced for AI cheating get normalized and repurposed for broader control (e.g. monitoring student attentiveness, etc.). Civil society groups and student unions in many countries are pushing back, emphasizing that education should not become a panoptic surveillance exercise (Avi, 2020; Scassa, 2023). This is a call for ODL institutions to find solutions that respect privacy – perhaps leveraging more trust-based and community proctoring models where possible, or technological solutions that are less intrusive (some are exploring “open book, open internet” exams graded differently).

4.4.5. Civilizational and Epistemological

Implications: At stake is what we count as evidence of learning. If models can draft plausible answers, assessment must weight process, originality of application, and ethical reasoning. Many high-AAPI policies already privilege iterative, contextualized work and reflective vivas, aligning integrity with human agency rather than tool detection. This reframing is not cosmetic; it redistributes power away from opaque classification toward dialogic judgment and transparent criteria, which is vital in multilingual, resource-constrained ODL systems (UNESCO, 2023; TEQSA, 2023).

4.4.6. Thought Leader Reflections: To bring voices to these points, consider a quote by James Zou (co-author of the detector bias study): *“Current detectors are clearly unreliable and easily gamed, which means we should be very cautious about using them as a solution to the AI cheating problem.”* (Stanford HAI, 2023) This encapsulates the folly of over-relying on tech when the tech itself is flawed – a direct plea for more holistic solutions. Another perspective, from Audrey Azoulay, UNESCO’s Director-General, emphasized that AI governance in education must keep the *“primary interest of learners”* at heart (North, 2023) – not the interest of appeasing rankings or political pressure. This means equity and inclusion should be front and center, as per UNESCO’s first guideline: *“promote inclusion, equity, linguistic and cultural diversity”* in AI use (North, 2023). If Global South institutions take that seriously, they will be wary of one-size-fits-all imported solutions and instead adapt AI to their multilingual, diverse student bodies – for instance, insisting that AI tools they use support local languages and are free of bias, and focusing on closing the digital divide (making sure all students have access to necessary technology and training) (North, 2023).

To sum up, the cross-disciplinary perspective underscores that AI in assessment is not just an IT or academic issue; it’s a sociotechnical and ethical governance

challenge. Getting it right is crucial for educational equity and the decolonization of knowledge – ensuring that the AI revolution in education does not simply impose a new form of Western hegemony or exacerbate inequality, but rather becomes an opportunity to re-imagine assessment in a way that is more fair, creative, and aligned with human development goals. Our analysis provides evidence and examples of both pitfalls and promising practices on this front. In the final section, we will synthesize these insights into concrete recommendations and a forward-looking roadmap.

4. Conclusion and Recommendations

5.1. Conclusion: Across 40 ODL providers, design-oriented, transparent policies are associated with healthier integrity signals than surveillance-heavy regimes. The AAPI makes this pattern visible and actionable. The policy choice is not between naïveté and control, but between opaque policing and deliberate design that treats integrity as a property of assessment ecosystems. This is consonant with mission-driven learning principles that prioritize formative process and ethical agency over product policing (Sangwa & Mutabazi, 2025a). For the Global South, the path to credibility runs through assessment redesign, disclosure and literacy, and enforceable rights protections. Institutions should publish safeguards as carefully as they publish sanctions, and vendors should be held to verifiable audit standards. Integrity in the AI era will be secured by what and how we ask students to do, not by how intrusively we watch them do it. The index gives a tractable pathway for incremental adoption through blueprint-level changes, authentic task design, scaffolded disclosure, supportive AI literacy, and procedural justice. In this configuration, proctoring is a specific tool used sparingly where learning outcomes warrant it, not a default posture that normalizes suspicion. The pilot scale and cross-sectional design necessarily caution against causal over-reach; nonetheless, convergent signals across our small-N sample and recent multi-modal invigilation studies (Alguacil et al., 2024) suggest that design-first governance, buttressed by proportional surveillance, can be both feasible and normatively superior. Future research should test AAPI’s predictive validity longitudinally and extend the index to programmatic accreditation settings. In sum, credible assessment under AI is less a technology race than a policy architecture: institutions that learn to design for integrity, disclose with clarity, and govern with fairness will remain trusted.

5.2. Recommendations: Design integrity, do not police it into being. First, redesign assessments toward authentic, contextual tasks that surface process and judgment, blending low-stakes iteration with selective vivas. Second, codify clear disclosure norms for permissible AI

assistance and teach AI literacy as a learning outcome. Third, adopt a proportional, multi-pronged invigilation protocol that couples (a) light-touch electronic analytics for anomaly detection, (b) dual-camera or 360-degree room-view systems calibrated to minimise privacy intrusions, and (c) trained human invigilators empowered to override algorithmic flags (van den Berg, 2025). This triadic arrangement reconciles deterrence with due-process safeguards and has been shown to outperform single-mode solutions in both deterring misconduct and sustaining student trust. Fourth, institute procedural safeguards: human review of detector flags, accessible appeals, and explicit non-punitive first-offense education where appropriate. Fifth, enact equity protections by acknowledging detector limitations, offering bandwidth-sensitive alternatives to high-stakes proctoring, and supporting multilingual pathways. Sixth, publish privacy impact assessments for any surveillance technology and retire tools that lack independent accuracy and bias audits. Seventh, invest in staff capacity to design and review assessments at scale, including communities of practice and shared open resources. Finally, align with credible external frameworks (e.g., UNESCO guidance; TEQSA's assessment reform principles) to promote consistency without importing black-box controls, and the World Economic Forum's Presidio Recommendations on responsible generative AI (WEF, 2023). Vendors should meet institutions where they are: privacy-by-design, low-bandwidth modes, transparent model reports, and independent audits are prerequisites, not add-ons (Buolamwini & Gebru, 2018; UNESCO, 2023; TEQSA, 2023).

5.3. Limitations: This research, being a secondary analysis, is limited by the availability and quality of public data on academic misconduct and policy details. A further constraint is sample size: forty institutions represent less than ten percent of the global ODL landscape, so findings must be treated as hypothesis-generating. Future work should replicate the AAPI with stratified samples and integrate longitudinal mixed-methods—combining quantitative indicators with ethnographic observations of invigilation practices—to strengthen external validity (Susnjak & McIntosh, 2024). Some correlations noted (e.g. between AAPI and flag rates) are suggestive but not proof of causation; unobserved factors (institutional culture, student demographics) could also play roles. The AAPI, while systematically constructed, inherently involves qualitative judgments; different weighting of dimensions could slightly alter scores. Furthermore, the rapid pace of change means policies and technologies are moving targets – our “snapshot” may be outdated quickly as new strategies emerge. Finally, our focus on ODL and Global South was broad; in-depth case studies of each institution were beyond scope, so contextual nuances might be glossed over. Despite

these limitations, the convergence of evidence from multiple sources lends confidence to our broad conclusions.

Conflicts of Interest The authors affirm that they hold no personal or institutional interests that could be construed as influencing the research or its interpretation.

Data Availability All original analyzes, interpretive frameworks, and illustrative excerpts are fully integrated into the manuscript. Any supplementary materials or clarifications that might assist replication or further investigation are available upon reasonable request from the corresponding author.

Research Ethics Because the inquiry relied exclusively on publicly available secondary sources, it involved no interaction with human participants or collection of identifiable personal data. Consequently, institutional ethics approval was not required under prevailing guidelines for social-science research. The study nonetheless adhered to accepted norms of responsible scholarship, including transparent sourcing, careful stewardship of data, and respect for intellectual property.

Funding Statement The investigation was undertaken solely with the authors' own time, expertise, and resources; it did not receive financial support from any external source.

Author	Contributions
(CRedit)	Conceptualization, methodology, data curation, formal analysis, visualization, original-draft preparation, and project administration were led by S. Sangwa. Investigation and critical revision of the text were carried out by S. Sangwa and P. Mutabazi. Supervision was provided by P. Mutabazi. Both authors satisfy Taylor & Francis authorship guidelines and have approved the final manuscript.

Use of Artificial Intelligence The authors employed the OpenAI GPT-5 model (accessed 1 September 2025) exclusively to enhance grammar, refine wording, and accelerate the discovery of publicly available scholarly sources related to the study. A typical prompt read: “*Improve the following paragraph for clarity and suggest peer-reviewed studies published since 2020 on this topic.*” The system generated no original concepts, data, analyzes, or figures; every AI-suggested edit or reference was independently reviewed for accuracy and appropriateness. The authors assume full responsibility for the manuscript's content.

Appendices

Appendix A. AAPI Codebook, Scoring, and Validation Details [Link, see also: <https://doi.org/10.17605/OSF.IO/VSC2F>]

References

- Alguacil, M., Herranz-Zarzoso, N., Pernías, J. C., & Sabater-Grande, G. (2024). Academic dishonesty and monitoring in online exams: A randomized field experiment. *Journal of Computing in Higher Education*, 36(4), 835–851. <https://doi.org/10.1007/s12528-023-09378-x>
- Avi Asher-Schapiro (2020, Nov 11). 'Unfair surveillance'? Online exam software sparks global student revolt. Thomson Reuters Foundation News. <https://news.trust.org/item/20201110125959-i5kmg>
- Bhabha, H.K. (1994). *The Location of Culture* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203820551>
- Bhaskar, R. (2008). *A Realist Theory of Science* (1st ed.). Routledge. <https://doi.org/10.4324/9780203090732>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency* (pp. 77-91). PMLR. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Center for Democracy & Technology (CDT). (2022). Courts recognize constitutional harms from invasive school surveillance. <https://cdt.org/insights/courts-recognize-constitutional-harms-from-invasive-school-surveillance/>
- Citizen Reporter. (2024, April 18). UNISA investigates thousands of students for dishonesty and plagiarism. *Rekord / The Citizen*. <https://www.citizen.co.za/rekord/news-headlines/2024/04/18/unisa-investigates-thousands-of-students-for-dishonesty-and-plagiarism/>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. <https://utstat.utoronto.ca/brunner/oldclass/378f16/readings/CohenPower.pdf>
- Coley, M. (2023, August 16). Guidance on AI detection and why we're disabling Turnitin's AI detector. Vanderbilt University. <https://www.vanderbilt.edu/brightspace/2023/08/16/guidance-on-ai-detection-and-why-were-disabling-turnitins-ai-detector/>
- Couldry, N., & Mejias, U. (2019). *The costs of connection: How data is colonising human life and appropriating it for capitalism*. Stanford University Press. <https://www.sup.org/books/sociology/costs-connection>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160. <https://doi.org/10.2307/2095101>
- Dawson, P., Bearman, M., & Dollinger, M. (2024). Validity matters more than cheating: Re-centering assessment design in the generative AI era. *Assessment & Evaluation in Higher Education*, 49(7), 945–960. <https://doi.org/10.1080/02602938.2024.2386662>
- Drozowski, M. J. (2025, July 8). Testing Turnitin's new AI detector: How accurate is it? *BestColleges*. <https://www.bestcolleges.com/news/analysis/testing-turnitin-new-ai-detector/>
- Friedman, B., Kahn Jr., P. H., & Borning, A. (2002). *Value sensitive design: Theory and methods* (Technical Report No. 02-12-01). University of Washington. <https://doi.org/10.21236/ada457085>
- Gignac, G. E., & Szodorai, E. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, 102, 74–78. <https://doi.org/10.1016/j.paid.2016.06.069>
- Hughes, J. M. C., & McCabe, D. L. (2006). Understanding Academic Misconduct. *Canadian Journal of Higher Education*, 36(1), 49–63. <https://doi.org/10.47678/cjhe.v36i1.183525>
- Human Rights Watch. (2022). “How dare they peep into my private life?” Children’s rights violations by governments that endorsed online learning during the Covid-19 pandemic. <https://www.hrw.org/report/2022/05/25/how-dare-t hey-peep-my-private-life/childrens-rights-violation s-governments>
- International Certification Accreditation Council (ICAC). (2023, April 14). Remote proctoring surges during Covid. Retrieved September 26, 2025, from <https://www.icacnet.org/remote-proctoring-surges-d uring-covid/>
- International Telecommunication Union. (2024). Global ICT indicators database. <https://www.itu.int/en/ITU-D/Statistics/Pages/publications/wtid.aspx>
- Griffey, A. (2024). Legal implications of remote exam proctoring: Extending federal law to protect student privacy. *Case Western Reserve Law Review*, 74(3), 791–836. <https://scholarlycommons.law.case.edu/caselrev/vol74/iss3/6>
- Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2024). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 5, 100118. <https://arxiv.org/abs/2405.11800>
- Lancaster, T. (2023). Academic Integrity for Computer Science Instructors. In: Carter, J., O'Grady, M., Rosen, C. (eds) *Higher Education Computer Science*. Springer, Cham. https://doi.org/10.1007/978-3-031-29386-3_6
- Lee, J. & Fanguy M. (2022). Online exam proctoring technologies: Educational innovation or deterioration? *British Journal of Educational Technology*, 53(6), 1533–1549. <https://doi.org/10.1111/bjet.13182>
- Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns* (New York, N.Y.), 4(7), 100779. <https://doi.org/10.1016/j.patter.2023.100779>

- Liu, D. Y.T., & Bates, S. (2025, January). Generative AI in higher education: Current practices and ways forward. Association of Pacific Rim Universities. Retrieved from https://www.apru.org/wp-content/uploads/2025/01/APRU-Generative-AI-in-Higher-Education-Whitepaper_Jan-2025.pdf.
- Louisiana State University Alexandria (LSUA). (2023, September 12). Proctorio – A guide and best practices for students. <https://helpdesk.lsua.edu/support/solutions/articles/48001166676-proctorio-a-guide-and-best-practices-for-students>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). Qualitative data analysis: A methods sourcebook (3rd ed.). SAGE. <https://collegepublishing.sagepub.com/products/qualitative-data-analysis-4-246128>
- Mochizuki, Y., Bruillard, E., & Bryan, A. (2025). The ethics of AI or techno-solutionism? UNESCO's policy guidance on AI in education. *British Journal of Sociology of Education*, 1–22. <https://doi.org/10.1080/01425692.2025.2502808>
- North, Madeleine. (2023, September 27). Generative AI has disrupted education. Here's how it can be used for good – UNESCO. World Economic Forum. Retrieved from <https://www.weforum.org/stories/2023/09/generative-ai-education-unesco/>
- Nyaaba, M., Wright, A., & Choi, G. L. (2024). Generative AI and digital neocolonialism in global education: Towards an equitable framework (arXiv:2406.02966v3). <http://arxiv.org/pdf/2406.02966v3>
- OECD (2023), OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem, OECD Publishing, Paris, <https://doi.org/10.1787/c74f03de-en>
- ProctorU. (2022). ProctorU pricing plan [PDF]. Scottsdale Community College. <https://www.scottsdalecc.edu/sites/default/files/inline-files/proctoru-pricing-plan-2022XC0623.pdf>
- Ramírez Castañeda, V. (2025, January 21). The creation of bad students: AI detection for non-native English speakers. D-Lab, UC Berkeley. <https://dlab.berkeley.edu/news/creation-bad-students-ai-detection-non-native-english-speakers>
- Salami, A. O. (2024). Artificial intelligence, digital colonialism, and the implications for Africa's future development. *Data & Policy*, 6, e67. doi:10.1017/dap.2024.75
- Sample, I. (2023, July 10). Programs to detect AI discriminate against non-native English speakers, shows study. *The Guardian*. <https://www.theguardian.com/technology/2023/jul/10/programs-to-detect-ai-discriminate-against-non-native-english-speakers-shows-study>
- Sangwa, S., & Mutabazi, P. (2025a). Mission-Driven Learning Theory: Ordering Knowledge and Competence to Life Mission. Preprints. <https://doi.org/10.20944/preprints202509.0906.v1>
- Sangwa, S., & Mutabazi, P. (2025b). I-Powered Multilingual Degrees in Higher Education: Postcolonial Insights on Linguistic Justice and Disruptive Innovation. In *Open Journal of AI, Ethics & Society (OJ-AES)* (Vol. 1, Number 1). Open Christian Press. <https://doi.org/10.5281/zenodo.17072137>
- Sangwa, S., Ngobi, D., Ekosse, E., & Mutabazi, P. (2025). AI governance in African higher education: Status, challenges, and a future-proof policy framework. *Artificial Intelligence and Education*, 1(1), Article 2054. <https://doi.org/10.62617/aje2054>
- Sangwa, S., Nsabiyumva, S., Ngobi, D., & Mutabazi, P. (2025). Ethical AI integration in African higher education: Enhancing research supervision, grant discovery, and proposal writing without compromising academic integrity. *International Journal of Human Research and Social Science Studies*, 2(6), 402–410. <https://doi.org/10.55677/ijhrss/08-2025-Vol0216>
- Scassa, T. (2023). The Surveillant University: Remote Proctoring, AI, and Human Rights. *Canadian Journal of Comparative and Contemporary Law*, 9, 301–342. <https://www.cjcl.ca/wp-content/uploads/2022/10-Scassa.pdf>
- Singh, M. (2023, October). Maintaining the integrity of the South African university: The impact of ChatGPT on plagiarism and scholarly writing. *South African Journal of Higher Education*, 37(5), 203–220. <https://doi.org/10.20853/37-5-5941>
- Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The End of Online Exam Integrity? *Education Sciences*, 14(6), 656. <https://doi.org/10.3390/educsci14060656>
- Sustainability Directory. (2025). Postcolonial AI Critique. Sustainability Directory. <https://climate.sustainability-directory.com/term/postcolonial-ai-critique/>
- Stanford HAI. (2023, May 15). AI Detectors Biased Against Non-Native English Writers. Retrieved from <https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers>.
- Stanford HAI. (2024). 2024 AI Index Report (Chapter 4: Technical Performance and Bias). Stanford Institute for Human-Centered AI. <https://aiindex.stanford.edu/report/>
- Teigaga, E. (2025). Addressing and mitigating examination irregularities: The role of invigilators. *International Journal of Multidisciplinary Research and Development*, 12(4), 35–38. <https://www.allsubjectjournal.com/assets/archives/2025/vol12issue4/12073.pdf>
- TEQSA. (2023). Advice to higher-education providers: Use of generative AI in assessment and academic integrity. Tertiary Education Quality and Standards Agency. <https://www.teqsa.gov.au/sites/default/files/2023-06/generative-ai-teqsa-guidance.pdf>
- The Cheat Sheet. (2024). 5,000 students accused of plagiarism at University of South Africa. [blog]. <https://thecheatsheet.substack.com/p/5000-students-accused-of-plagiarism>
- Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall

- method of coal-getting. *Human Relations*, 4(1), 3–38.
<https://doi.org/10.1177/001872675100400101>
- Turnitin. (2023, July 25). Turnitin AI detection feature reviews more than 65 million papers. <https://www.turnitin.co.uk/press/turnitin-ai-detection-feature-reviews-more-than-65-million-papers>
- Turnitin. (2024, April 5). Turnitin marks one-year anniversary of its AI writing detector with new data insights on student use of generative AI. <https://www.turnitin.com/press/turnitin-first-anniversary-ai-writing-detector>
- UNICEF. (2020, August). COVID-19: Are children able to continue learning? A global analysis of the potential reach of remote learning policies. UNICEF. <https://data.unicef.org/resources/remote-learning-reachability-factsheet/>
- UNICEF. (2021). Data for action on remote learning during COVID-19: Where do children have internet access at home? <https://data.unicef.org/data-for-action/where-do-children-have-internet-access-at-home-insights-from-household-survey-data/>
- UNISA (University of South Africa) (2024, July 5). Media Statement: Unisa clarifies stance on its zero-tolerance against academic misconduct and use of AI tools. https://www.unisa.ac.za/static/corporate_web/Content/News%20&%20Media/Media%20releases/documents/Media%20Statement%20-%20Unisa%20clarifies%20stance%20on%20its%20zero-tolerance%20on%20academic%20misconduct%20and%20use%20of%20AI%20tools.pdf
- University of British Columbia (UBC). (2023). Designing for academic integrity. Centre for Teaching and Learning. <https://ctl.ok.ubc.ca/teaching/course-design/assessment/designing-for-academic-integrity/>
- UNESCO. (2023). Guidance for generative AI in education and research. https://unesco.org.uk/site/assets/files/10375/guidance_for_generative_ai_in_education_and_research.pdf
- UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence: Key facts (SHS/2023/PI/H/1). UNESCO. https://unesco.org.uk/site/assets/files/14137/unesco_recommendation_on_the_ethics_of_artificial_intelligence_-_key_facts.pdf
- van den Berg, L. (2025). Re-thinking invigilation implementation: A mixed-method approach to student perceptions and system usability for digital assessment adoption. *International Journal for Educational Integrity*, 21, Article 10. <https://doi.org/10.1007/s40979-025-00183-w>
- Wang, T. (2023). Navigating Generative AI (ChatGPT) in Higher Education: Opportunities and Challenges. In: Anutariya, C., Liu, D., Kinshuk, Tlili, A., Yang, J., Chang, M. (eds) *Smart Learning for A Sustainable Society*. ICSLE 2023. Lecture Notes in Educational Technology. Springer, Singapore. https://doi.org/10.1007/978-981-99-5961-7_28
- Word-Spinner. (2025). How much AI is allowed in Turnitin? Word-Spinner. <https://word-spinner.com/blog/how-much-ai-is-allowed-in-turnitin/>
- World Bank. (2024). World development indicators (database). Retrieved from: <https://databank.worldbank.org/source/world-development-indicators>
- World Economic Forum. (2023, June 14). The Presidio recommendations on responsible generative AI. https://www3.weforum.org/docs/WEF_Presidio_Recommendations_on_Responsible_Generative_AI_2023.pdf
- Yeo, M. A. (2023). Academic integrity in the age of Artificial Intelligence (AI) authoring apps. *TESOL Journal*, 14(1), e716. <https://doi.org/10.1002/tesj.716>