Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Institutional Governance Authenticity and Student Responses to Artificial Intelligence Assessment Guidelines in Higher Education

Peter Tor Atorough

peter.atorough@dmu.ac.uk

De Montfort University

Sara Hajikazemi

Queen Mary University of London

Perspective

Keywords: academic integrity, procedural justice, neutralisation theory, reputational greenwashing, ethical legitimacy

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11016438/v1

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

Additional Declarations: No competing interests reported.

Abstract

Rapid proliferation of generative artificial intelligence (GenAI) has prompted universities worldwide to issue guidelines on student AI use in assessment. While existing research catalogues policy content and student misuse, little is known about how students perceive these guidelines and how such perceptions shape behaviour. This paper argues that current research treats guidelines as a binary intervention—present or absent—overlooking their psychological and institutional authenticity. Drawing on procedural justice and neutralisation theories, we contend that when guidelines are vague, invisible, unenforceable, and treated by academics as box-ticking exercises, they fail as integrity safeguards and may instead provide justifications for disingenuous AI use. We further argue that this pattern reflects institutional compulsions—legal, reputational, and competitive—that drive guideline issuance despite known limitations. We propose a research agenda centred on guideline authenticity and its effects on fairness perceptions, moral neutralisation, and compliance intentions, alongside perceptions of institutional motivation. A sequential mixed-methods design is outlined, drawing on UAE and UK contexts. The agenda aims to move institutional practice beyond symbolic policy-making toward genuinely embedded academic integrity cultures.

Introduction

Universities exercise authority not through force but through legitimacy – the belief on the part of students and staff that institutional rules are fair, justified, and enacted in good faith. When students regard institutional policies as clear, coherent, and sincerely held, they may be more likely to internalise them as ethical obligations. On the other hand, however, if rules are perceived as performative gestures – issued to satisfy external audiences while remaining practically unenforceable – the relationship between institution and individual may weaken. In this instance, compliance becomes a matter of calculation rather than conviction, potentially eroding the norms that academic integrity policies seek to uphold.

The advent of generative AI (GenAI) in higher education has forced a reckoning over how universities govern student behaviour. Now that GenAI can instantly generate passing-grade essays, code, and creative projects without triggering traditional plagiarism software (Corbin et al., 2025a; Gonsalves, 2024), universities have scrambled. The immediate and almost universal fix has been a wave of new guidelines, stretching from total bans to complex, tiered disclosure rules (Perkins et al., 2024; Luo, 2024). But this flood of paperwork begs a deeper question: do these rules actually influence students, or have they just become empty gestures that fail to change actual behaviour on the ground?

While plenty of studies track how students misuse AI, research into the policies themselves shows they are messy, reactive, and disconnected from the classroom (Luo, 2024). Corbin et al. (2025a) argue that relying on mere words – rules that students can easily bypass – creates an "enforcement illusion" that leaves the integrity of exams totally exposed. Frankly, “talk is cheap”.

What we often overlook is why universities feel so desperate to publish these guidelines, even when they know the rules are, in practice, of limited enforceability. We suggest that this panic stems from a "perceived governance authenticity" problem. Universities don't operate in a vacuum; they are constantly compelled by legal mandates, public scrutiny, and the need to emulate their peers. They have to prove they are maintaining standards, protect their brand to outsiders, and match whatever their competitors are doing (cf. QAA, 2023). This creates a massive incentive to look busy and project control, even when administrators know their new rules are completely unenforceable.

The above dynamic can plausibly be understood as a form of symbolic governance (what one might term reputational greenwashing in more critical framings): policy statements that signal institutional responsiveness while masking limited structural change (cf. Corbin et al., 2025a). From an institutional theory perspective, this is not necessarily strategic failure but rather a predictable response to competing pressures. Nevertheless, when the gap between policy and practice becomes visible to students, it diminishes their perception of institutional authority.

Educators and students see right through this. Caught between bureaucratic demands and actual classroom life, teachers usually know when official guidelines won't work on the ground (Shareen & Al Sayari, 2025; Jiang et al., 2025). So, they do the bare minimum. They stick the AI rules into the syllabus and leave it at that. Students notice this half-hearted effort, and they read between the lines, creating a domino effect of check-the-box mentality, which ultimately reshapes how students weigh their own ethical choices.

This paper builds a framework to test exactly how that domino effect plays out. We argue that relying on rigid AI policies has become a bureaucratic security blanket – it satisfies auditors and protects a university's reputation but does little to change real student behaviour (Harris, 2024; Castelló Sirvent et al., 2024). The real issue is the tone. Most AI policies feel like police manuals rather than educational guides (HEPI, 2026). By treating AI purely as a loss-prevention issue (Atorough & Donaldson, 2014), institutions focus heavily on catching cheaters while overlooking deeper values such as academic honour. In the end, this relentless focus on penalties burns out any internal desire a student has to do the right thing.

Through the lens of two primary questions, we explore these effects. We ask:

1. To what extent do perceived governance deficiencies – vagueness, invisibility, unenforceability, and minimal enactment – shape students' procedural justice evaluations, neutralisation justifications, and intentions to use AI disingenuously?
2. How do students and educators interpret the institutional drivers behind guideline issuance, and how do these perceptions shape their responses?

To address these questions, we develop a conceptual framework that integrates procedural justice theory (Colquitt, 2001; Colquitt et al., 2001) and neutralisation theory (Sykes & Matza, 1957), employing a

mixed methods research design that traces pathways from institutional motivation through guideline design and enactment to behavioural intention.

Specifically, we make three contributions: we conceptualise perceived governance authenticity as a determinant of students' responses to institutional policy; integrate procedural justice and neutralisation into a unified explanatory framework; and advance a multi level research agenda for understanding AI governance in higher education. In doing so, we move the conversation from AI assessment policy to its implications for the governance of higher education and student compliance behaviour. Accordingly, this paper is presented as a Perspective Article that develops a conceptual framework and future research agenda.

The Role of Institutional Legitimacy in Educational Governance

Before turning to the literature review, it is worth foregrounding the conceptual foundations of our argument. Educational institutions exercise authority not through force but through legitimacy: the belief on the part of students and staff that institutional rules are fair, justified, and enacted in good faith. This legitimacy is a resource that may be inherently fragile (Brennan & Naidoo, 2008; Macfarlane, 2004). When institutional pronouncements are perceived as insincere – issued to meet external regulatory demands or to manage reputation – the basis of rule following may shift from an internalised sense of obligation to an external calculus of risk and reward. In the language of moral psychology, students who would otherwise comply out of a sense of integrity may instead ask, "What can I get away with?" The guideline, rather than reinforcing ethical norms, may become an incentive for moral disengagement.

This points to a deeper truth: authenticity isn't a luxury. When a university puts out a rule without genuinely backing it up, that rule loses the moral weight needed to successfully change minds. It just looks like paperwork. On the flip side, things change when institutions actually sit down with students and faculty to co-create these boundaries. Visibly weaving these rules into daily teaching – and explaining the why behind them – proves the policy isn't just a shield to ward off lawsuits. It shows the institution cares about learning. This goes way beyond clever policy design. In organisational theory, legitimacy hinges on whether people think a system is fair. If they do, they follow the rules because they want to, not because they are terrified of getting caught.

Critical Literature Review and Conceptual Framework

From Guideline Content to Student Perception: Identifying the Gap

Right now, research on AI in assessment splits into a few distinct conversations. Mostly, scholars are looking at students – how quickly they have adopted these tools, their blurry definitions of what counts as cheating, and the messy line between using AI for study help versus outright shortcuts (Gonsalves, 2024). Another major focus is on how universities are scrambling to react. In a look at policies across twenty top-tier universities, Luo (2024) used Bacchi’s "What's the problem represented to be" framework

to show that institutions treat GenAI as a fire to be put out rather than a tool to build with. These policies rest on shaky assumptions about how students act and mistakenly assume we can accurately catch AI-generated text. That brings us to the tech itself: a mountain of evidence shows that AI detectors are deeply unreliable, prone to false positives, and introduce massive equity issues. A few researchers are trying to flip the script by redesigning exams entirely – focusing on building "evaluative judgment" so students learn to use AI critically (Bearman et al., 2024). But changing the actual structure of an exam is still rare. Most universities prefer the easier route: just writing another policy document.

Conspicuously thin is work that examines how students experience the policies designed to govern their behaviour. A handful of studies have explored awareness of general academic integrity policies, finding that students often do not read honour codes or view them as irrelevant (Ellis & Murdoch, 2024). In the AI context specifically, Gonsalves (2024) found that ambiguous guidelines and inconsistent enforcement were among the primary drivers of non-compliance with AI use declarations. Corbin et al. (2025b) reported that both students and staff described the boundary between acceptable and unacceptable AI use as "absurd," with students constructing "individually unique and often complex ethical models for AI use" in the absence of clear institutional guidance. This is significant oversight because, from a psychological perspective, a rule that is not perceived as real in practice may struggle to exert normative force. This is consistent with a broader body of work in higher education that emphasises institutional trust, pedagogical responsibility, and the legitimacy of educational authority as foundational to student behaviour.

Why Guidelines Proliferate: Institutional Pressures and Symbolic Governance

To understand why stakeholders may perceive guidelines as inauthentic, we must examine the institutional conditions that produce them. We propose three intersecting drivers, each of which may generate pressure to issue guidelines irrespective of practical enforceability.

Foremost, universities face legal and moral obligations to maintain academic standards. Quality assurance frameworks, such as those administered by the Quality Assurance Agency (QAA) in the UK and the Commission for Academic Accreditation (CAA) in the UAE, require institutions to demonstrate that assessment practices are valid, reliable, and secure. The arrival of GenAI represents a fundamental challenge to these assurances. Issuing a guideline constitutes a demonstrable institutional response – evidence of due diligence that can be presented to regulators, external examiners, and professional accrediting bodies. The moral obligation operates in parallel: universities hold a duty of care to certify that graduates have genuinely achieved the stated learning outcomes. A guideline, however unenforceable in practice, symbolises the institution's acknowledgement of this duty. The legal moral driver thus operates primarily at the level of institutional accountability to external bodies; the result is an incentive to document a response even when the operational response remains underdeveloped.

Secondly, universities confront reputational pressures. The perceived value of a degree depends on public confidence in the maintenance of academic standards. In a competitive higher education marketplace, any suggestion that an institution has failed to address AI related integrity threats carries

reputational risk. Every KPI is on the line here – rankings, recruitment, employment stats, and how employers view the university. Corbin et al. (2025a) nail this dynamic with their concept of the "enforcement illusion." They point out that these guidelines use all the scary, official words – bans, disclosures, penalties – but lack any real teeth or backend infrastructure to back them up. Still, this illusion keeps the public happy. It tells outsiders that the university has everything under control, even if the professors on the inside know it's just a performance. We call this symbolic governance – or, more bluntly, reputational greenwashing. The goal isn't to accuse universities of outright lying. Rather, it highlights an uncomfortable reality: these policies often exist to protect a school's brand rather than to actually help students learn.

And finally, universities operate within a field of competitive isomorphism (DiMaggio & Powell, 1983). When peer institutions issue AI guidelines, there is institutional pressure to follow suit – failure to do so risks appearing complacent, reactive, or indifferent to a widely recognised threat. Mimetic isomorphism – the imitation of peer institutions' structures and practices under conditions of uncertainty – is particularly relevant. The rapid and still evolving nature of GenAI means that no institution can claim definitive expertise; in this uncertainty, adopting similar policy responses to comparator institutions may become the safest strategic choice. The result may be a convergence of guideline forms, even when those forms are known to be practically limited. This competitive dynamic may create a self-reinforcing cycle: the more universities issue guidelines, the more other universities must do the same, regardless of whether the guidelines themselves are fit for purpose. For institutions, the question of relevance shifts from "what works?" to "what are others doing?" – a shift that, from a governance perspective, may substitute genuine integrity with conformity.

The Lecturer's Position: Mediating Between Mandate and Reality

Educators occupy a structurally ambivalent position within this institutional dynamic, acting simultaneously as the agents of institutional policy and the practitioners most directly exposed to its limitations. Tasked with implementing guidelines in their courses, they confront a daily gap between what the guidelines prescribe and what is practically enforceable or pedagogically desirable. When a lecturer judges a guideline to be unenforceable – for instance, because detecting AI use in formative reflective writing is acknowledged to be impossible – they face a choice: invest disproportionate effort in enforcement that yields uncertain returns or perform a minimal compliance gesture and redirect attention to more tractable aspects of teaching. The latter response, observable as minimal enactment, is not necessarily an act of indifference, but may reflect a rational allocation of limited professional resources in the face of an impossible mandate.

This dynamic carries considerable implications for perceptions of procedural justice. Xia and Wei (2026) demonstrated that implementation fidelity – the observable enactment of governance in courses – matters more than policy content for shaping student trust and transparency behaviours. When educators, as the visible face of institutional governance, signal through their behaviour that the guideline is not to be taken seriously, the procedural justice of the entire governance framework may be

undermined. Our framework positions the lecturer not as a neutral conduit of policy but as an active mediator whose own perceptions of governance authenticity – shaped by their understanding of the institutional motives behind it – influence how they enact it, which in turn shapes what students perceive. Research that excludes the lecturer perspective therefore misses a critical mechanism in the governance authenticity cascade.

The Limitations of Existing Behavioural Models

Most research on academic integrity leans heavily on the Theory of Planned Behaviour (TPB). The basic idea is that what students intend to do comes down to three things: their own attitude, peer pressure, and how much control they think they have. Gonsalves (2024) used this exact model to look at whether students declare their AI use, proving that vague rules, peer influence, and a fear of getting caught are the biggest hurdles. But while this model keeps things neat and simple, it doesn't quite work for what we are trying to uncover. It reduces an incredibly messy ethical choice to a basic math equation. In doing so, it completely misses the psychological gymnastics students go through to justify their choices and bypass their own moral boundaries. Critically, it does not account for the perceived legitimacy of the rule making process itself – only the perceived consequences and social pressures around the behaviour. Given our interest in why guidelines may fail, and in the institutional motivations and lecturer mediations that contribute to that failure, a theoretical lens is needed that foregrounds the psychology of rule following and rule breaking in direct response to institutional action.

Theoretical Foundations: Procedural Justice and Neutralisation

Building on recent empirical work that has begun to bridge the gap between institutional governance and student behaviour, we propose a dual theory framework that captures both the institutional signal (the guideline's perceived fairness) and the individual cognitive response (the moral justifications that enable disengagement). This encapsulates governance and student behaviour.

Xia and Wei (2026), in a four-wave longitudinal study of 739 students across 25 courses, demonstrated that implementation fidelity – the extent to which GenAI governance is enacted consistently and visibly in courses – was a stronger predictor of procedural justice perceptions than policy explicitness. Procedural justice, in turn, predicted student trust and, through trust, greater AI disclosure and less concealment. This finding is pivotal: it suggests that the way governance is carried out matters more than what governance documents say. In the proposed framework, we extend this insight by explicitly theorising the moral psychological pathway through which symbolic governance – rooted in institutional pressures and mediated by lecturer enactment – may not merely fail to promote transparency but actively license concealment and moral disengagement.

Procedural Justice Theory (Leventhal, 1980; Colquitt, 2001) holds that people's willingness to comply with rules is strongly influenced by their perceptions of the fairness of the process by which those rules are made and implemented. This relationship has been demonstrated in higher education contexts, where students' experiences of procedural fairness significantly shape their trust in and perceived

legitimacy of university authorities (Główczewski & Burdziej, 2023). Key dimensions of interest include voice (opportunity to participate in rule making), consistency (uniform application across cases and people), accuracy (based on reliable information), and ethicality (upholding dignity and respect). When these dimensions are satisfied, users are more likely to internalise rules as legitimate, and compliance becomes a matter of personal obligation rather than mere fear of punishment. When they are violated, legitimacy may erode, and rule breaking can feel justified. Xia and Wei (2026) have provided recent empirical evidence suggesting that this mechanism operates in the context of GenAI governance.

Applied to AI assessment guidelines, procedural justice provides a powerful diagnostic lens. A guideline developed without student input (low voice), applied differently by different educators (low consistency), based on an inaccurate understanding of how students use AI (low accuracy), and communicated with an implication that students are inherently untrustworthy (low ethicality) is likely to be perceived as procedurally unjust. We extend this framework by positing that students are also sensitive to the perceived commitment of institutional governance. When academics themselves treat the guideline as a box ticking exercise – pasting it into a course outline without discussion, failing to enforce compliance with it, or never referencing it in teaching – this may constitute a violation of procedural justice. The institution signals that the rule does not really matter, and students may learn to discount its moral force accordingly. The addition of the institutional motivation lens further deepens this analysis. We envisage that when students perceive that the guideline was issued primarily for reputational protection or regulatory compliance, rather than genuine pedagogical commitment, the procedural justice violation is compounded. The rule is not merely poorly implemented; it is seen as lacking sincerity from its inception, potentially compromising the normative legitimacy on which voluntary compliance depends.

The foregoing argument resonates with Corbin et al.'s (2025a) distinction between discursive and structural changes in assessment. Discursive changes – instructions, declarations, traffic light classifications – rely entirely on "student compliance with instructions, a reliance which opens both assessments to vulnerabilities as well as educational institutions to reputational risk." We argue that this vulnerability is not merely practical but deeply psychological: when students perceive that the institution itself treats its own rules as performative – as symbolic governance – the moral obligation to comply may be weakened.

Neutralisation Theory (Sykes & Matza, 1957) explains how individuals who generally hold conventional moral beliefs can engage in deviant behaviour by temporarily neutralising the moral bind imposed by a rule. Techniques include denial of responsibility, denial of injury, condemnation of the condemners, and appeal to higher loyalties. In the context of academic integrity, neutralisations such as "everyone else is doing it" or "it's not a big deal because I'm still learning" have been robustly linked to self-reported cheating. Flutti et al. (2026), in a study of 205 Italian university students, found a strong positive correlation between neutralisation techniques and academic dishonesty, with denial of responsibility exhibiting the strongest association with dishonest behaviour. Additional support is provided by Ellis and Murdoch (2024), who proposed an "educational integrity enforcement pyramid" that implicitly acknowledges the role of perceived legitimacy of the institution's regulations in shaping student

behaviour, arguing that responses to cheating must be proportionate and grounded in educational principles rather than purely punitive measures.

Critically, neutralisation theory does not treat these justifications as purely pathological. They can be seen as rational adaptations to a perceived unjust or incoherent system. We argue that perceived inauthentic guidelines – and the institutional motivations that may produce them – could provide rich raw material for neutralisations. A vague guideline invites denial of responsibility ("How was I supposed to know?"). An unenforceable guideline invites condemnation of the condemners ("The university doesn't really care, so why should I?"). A guideline that is treated as mere box ticking reinforces the perception that the institution itself is acting instrumentally, thereby licensing the student to adopt a similarly calculating posture. Most powerfully, when students perceive the guideline as symbolic governance – a public performance rather than a genuine commitment – it gives them the ultimate excuse to circumvent the guidelines; a sort of meta-justification. They don't just view the university as a bad referee; they see the institution itself as acting in bad faith. A student who thinks, "the university only wrote this rule to cover its own back," instantly strips the entire policy of its moral weight. Any personal obligation to follow the rules evaporates. Ironically, instead of stopping misconduct, the policy itself becomes the perfect excuse to do it. The moment students spot the institutional hypocrisy behind a rule, their own moral guardrails come down.

The Conceptual Framework: An Extended Perceived Governance Authenticity Pathway

Synthesising these arguments, we propose a conceptual model that forms the core of the research agenda. The model extends the core pathway by incorporating the institutional and lecturer dimensions:

Before presenting the model, it is important to clarify the relationship between three related but distinct constructs:

- **Perceived governance authenticity** refers to students' perceptions that institutional AI guidelines are clear, visible, enacted by educators, and enforceable. It captures perceptions of institutional sincerity and the extent to which governance actions align with institutional claims. Authenticity is about the perceived genuineness of governance.
- **Procedural justice** refers to perceptions of fairness in how rules are made and implemented – whether students have voice, whether rules are applied consistently, and whether the process respects their dignity. Procedural justice is about fairness of process.
- **Legitimacy** refers to the broader acceptance of institutional authority – the belief that the institution has the right to govern and that its rules ought to be followed. Legitimacy is about acceptance of authority.

These constructs are conceptually related but distinct. A guideline could be perceived as procedurally fair (e.g., students were consulted) yet inauthentic (e.g., staff do not enforce it). Or it could be perceived

as authentic (e.g., staff take it seriously) yet procedurally unfair (e.g., students had no voice in its development). Legitimacy represents the overall outcome: when governance is perceived as both authentic and procedurally just, institutional authority is more likely to be accepted as legitimate.

Our model uses perceived governance authenticity as a formative higher order construct comprising three dimensions derived from the earlier critique and substantiated by recent empirical findings:

- **Clarity:** The degree to which the guideline is specific, comprehensible, and applicable to concrete assessment tasks. Corbin et al. (2025b) recommend moving beyond simple prohibitions or permissions toward guidance that addresses the feasibility of enforcement, the preservation of authentic learning, and the emotional wellbeing of teachers and students.
- **Visibility:** The extent to which the guideline is prominently and repeatedly communicated at the points of assessment decision making. Gonsalves (2024) found that even when declarations were mandatory, poor visibility contributed to 74% non-compliance.
- **Enforceability:** The belief that violations are likely to be detected and met with meaningful, consistently applied consequences. Corbin et al. (2025a) argue that discursive approaches create an "enforcement illusion," borrowing the language of socially familiar structural systems while lacking the actual enforcement capabilities of those systems.

We also specify mediation and exploratory moderation paths. First, **Perceived Lecturer Enactment** is the mediating component of the model: Students' observation that their lecturers and tutors take the guideline seriously, discuss it meaningfully, and apply it consistently. Xia and Wei (2026) demonstrate that implementation fidelity – the observable enactment of governance – matters more than the written policy itself. Second, **Perceived Institutional Motivation** is the extent to which students and tutors perceive the guideline as driven by genuine pedagogical commitment rather than by reputational protection, regulatory compliance, or competitive mimicry. This dimension captures perceptions of the institutional sincerity underlying the guideline and is hypothesised to moderate the relationship between the other authenticity dimensions, enactment and procedural justice: when perceived motivation is instrumental, even a relatively clear and visible guideline may be discounted.

Proposed Investigation Methods and Methodology

The current paper provides a conceptual perspective that requires future validation through empirical data. To empirically test and refine this conceptual model, we propose a sequential mixed methods research programme comprising two primary phases. A proposed future programme of research would collect data from university students in the United Arab Emirates and the United Kingdom.

These contexts were purposively considered to maximise variation in institutional guideline design, regulatory traditions, and student demographics. The UAE higher education sector operates within a distinctive regulatory environment characterised by strong centralised quality assurance (via the CAA), a high proportion of transnational education partnerships, and intense inter institutional competition; the

UK offers a more established and decentralised context. “Country” will be treated as a control variable in primary analyses; direct cross-national comparison does not constitute a research objective. Should the proposed research programme proceed to empirical implementation, ethical approval will be obtained from the relevant institutional ethics committees and informed consent will be obtained from all participants before data collection commences.

Phase 1: Qualitative Exploration

Purpose. To explore how students and academics perceive AI guidelines, the institutional motivations behind them, and the neutralisation narratives that students construct in response. This phase will also generate ecologically valid items for the subsequent quantitative instrument.

Design. Semi structured interviews and focus groups.

Proposed Participants. Purposive sampling of students across disciplines (humanities, STEM, creative arts) and year levels, alongside academics (module coordinators, programme leaders) from at least two institutions in the UAE and two in the UK. This multi perspective, multi-site design mirrors the approach successfully employed by Corbin et al. (2025b), who interviewed 19 students and 12 staff to uncover the "absurd line" between acceptable and unacceptable AI use. By including academics, we directly access the minimal enactment dynamic that Gonsalves (2024) identified as a driver of inconsistent enforcement, as well as educators' own perceptions of institutional objectives and how these perceptions shape their implementation behaviour.

Data collection. Participants will be presented with anonymised excerpts of real institutional AI guidelines and concrete assessment scenarios. Interview prompts will target the four perceived governance authenticity dimensions (clarity, visibility, enforceability, and lecturer enactment), and a dedicated set of prompts will explore perceived institutional motivation from both student and lecturer perspectives. Neutralisation prompts will probe for justifications linked to institutional flaws and perceived institutional insincerity.

Analysis. A hybrid deductive-inductive thematic approach will be applied in the data analysis. To start, we will build an initial coding framework around clearly defined core pillars: the dimensions of procedural justice, perceived authenticity, and Sykes and Matza’s classic neutralisation techniques. We will also code specifically for how students read institutional motives – tracking whether they view a policy as a genuine legal-moral duty, an act of peer-copying (competitive isomorphism), or mere symbolic governance. From there, we will map out how these elements interact. This thematic map will trace the exact lines connecting a university's perceived motives, its failures in authenticity, how lecturers handle the rules in class, and the specific excuses students use to justify cheating.

Phase 2: Quantitative Survey and Model Testing

Purpose. To test the hypothesised structural pathways from perceived governance authenticity through procedural justice and neutralisation to behavioural intentions, and to examine how perceptions of

institutional motivation moderate these pathways.

Design. Cross sectional, self-report online survey.

Proposed Participants. A sample of university students will be recruited across multiple institutions in the UAE and the UK to ensure variance in guideline exposure. The stratification strategy will follow discipline, year, and institution type, while a parallel survey of educators will capture the instructor perspective on the same constructs, enabling multi-level triangulation.

Instruments.

- Perceived Governance Authenticity Scale: Newly developed based on Phase 1 findings and theoretical dimensions, capturing clarity, visibility, enforceability, and lecturer enactment as a formative latent construct.
- Perceived Institutional Motivation Scale: Newly developed to capture Legal Moral, Reputational, Competitive, and Pedagogical motivation dimensions.
- Procedural Justice Perceptions: Adapted from Colquitt's (2001) organisational justice scale, contextualised to AI guidelines, and informed by Xia and Wei's (2026) operationalisation.
- Neutralisation Acceptance: Adapted scales from Flutti et al. (2026), with items reframed to capture institutionally directed neutralisations and moral disengagement.
- Behavioural Intention: Following established scenario-based methods employed by Gonsalves (2024), 2–3 assessment vignettes will be presented, varying in ambiguity. Participants rate on a 0–100% likelihood scale their intention to (a) comply fully and declare use, and (b) use AI beyond the guideline's limits without declaring.
- Covariates: Demographics, AI literacy, prior academic integrity training, actual guideline readership, and a social desirability scale.

Analysis and Hypotheses

The hypothesised structural model will be tested using structural equation modelling (SEM), with the following key hypotheses:

- **H1:** Perceived governance authenticity (comprising clarity, visibility, enforceability, and perceived lecturer enactment) is positively associated with students' perceptions of procedural justice.
- **H2:** Perceived lecturer enactment partially mediates the relationship between institutional governance characteristics (clarity, visibility, enforceability) and student procedural justice evaluations. This pathway captures the mechanism by which institutional policy features are translated into student experiences through educator behaviour.
- **H3**: Procedural justice is negatively associated with the acceptance of moral neutralisation techniques.

- **H4:** Neutralisation acceptance is positively associated with intentions to engage in disingenuous GenAI use (and negatively associated with compliance intentions).
- **H5:** Procedural justice and neutralisation acceptance serially mediate the relationship between perceived governance authenticity and behavioural intentions.
- **H6a** (Moderation): Perceived instrumental institutional motivation moderates the positive relationship between governance authenticity and lecturer enactment, such that higher perceived instrumental motivation weakens the relationship.
- **H6b** (Moderation): Perceived instrumental institutional motivation moderates the positive relationship between governance authenticity and procedural justice, such that higher perceived instrumental motivation weakens the relationship.

Figure 2 depicts the hypothesized paths and relationships, including moderation, direct and indirect relationships.

Conclusion: A Research Agenda for Understanding Governance Perceptions and Integrity Cultures

This paper has argued that the predominant institutional response to AI in assessment – the issuance of guidelines – has been implemented with insufficient attention to how the very students perceive those guidelines they aim to govern, and to the institutional pressures that may produce guidelines ill-suited to genuine behavioural influence. Recent scholarship provides converging evidence for this concern. Corbin et al. (2025a) have demonstrated that discursive rule statements may create an "enforcement illusion" rather than genuine assessment security. Gonsalves (2024) has shown that ambiguous guidelines and inconsistent enforcement drive widespread noncompliance with AI declarations. Xia and Wei (2026) have provided direct evidence that procedural justice and implementation fidelity – not merely the existence of policy – shape students' transparency behaviours. Corbin et al. (2025b) have revealed that students and staff alike find the boundaries drawn by current policies "absurd." Meanwhile, Bearman et al. (2024) have shown that a pedagogically richer alternative – developing students' evaluative judgement – is available yet remains overshadowed by the regulatory reflex to issue guidelines.

We expand on these ideas by mapping out the hidden pressures that keep these superficial guidelines alive. Specifically, we examine legal mandates, peer imitation, and symbolic branding – what we call reputational greenwashing. These forces push universities to keep using guidelines as a proxy for actual management, even when they know the rules won't work. Crucially, we also place the lecturer at the centre of this story. How teachers read the university's real motives completely shapes whether a policy feels authentic in the classroom. By focusing on governance authenticity, institutional motives, fairness, and student excuses, this research agenda shifts the entire conversation. We argue that AI cheating isn't just a failure of student morality. Instead, it is a direct consequence of flawed institutional design, organisational panic, and hollow policy enforcement. The mixed methods programme outlined will generate granular, theoretically grounded evidence on whether and how guidelines influence students'

moral reasoning, whether perceptions of institutional motivation colour the entire pathway, and whether, in their current form, guidelines serve as little more than symbolic proxies for genuine integrity reform.

The practical implications are significant. If the hypothesised pathways are supported, universities would need to move beyond simply updating policy documents to implementing governance perceived as authentic: cocreated with students and educators, visibly embedded in teaching, and consistently enforced. Further, the outcomes will provide evidence to support assessment redesign that reduces the incentive for misuse, and – crucially – is motivated by genuine pedagogical commitment that is transparently communicated to all stakeholders. This is an important consideration, given that institutions must also reckon candidly with the reputational and competitive pressures that shape their policy choices, recognising that guidelines issued primarily for external audiences may risk undermining the internal legitimacy on which compliance depends. As Corbin et al. (2025a) conclude, "talk is cheap" – the challenge is to build integrity into the architecture of assessment itself, and to ensure that the institutional voice speaking that talk is perceived as sincere. Only then can guidelines function as true anchors of a learning centred integrity culture, rather than as box ticking exercises that may inadvertently enable the behaviour they purport to prevent. This research agenda offers a roadmap for investigating that essential governance transformation.

Declarations

Disclosure statement:

No potential conflict of interest was reported by the authors.

Ethical approval

Not applicable. This article is a conceptual perspective paper and does not report research involving human participants, human data, or human tissue. No empirical data were collected or analysed. Ethical approval will be sought from the appropriate institutional ethics committee before the commencement of any future empirical research described within the proposed research agenda.

Consent to participate

Not applicable. No human participants were involved in this study.

Consent to publish

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Funding:

This research has not currently received any specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

The authors list the following CRediT contributions: Author 1 P.T.A: conceptualisation, methodology, visualisation, writing - initial. Author 2 S.H: Conceptualisation, synthesis, writing - revision and editing.

Data Availability

Not applicable.

References

1. Alsharefeen R, Al Sayari N. Examining academic integrity policy and practice in the era of AI: A case study of faculty perspectives. Front Educ. 2025;10., Article 1621743. https://doi.org/10.3389/feduc.2025.1621743.
2. Atorough P, Donaldson B. The relationship between regulatory focus and online shopping – perceived risk, affect, and consumers' response to online marketing. Int J Internet Mark Advertising. 2012;7(4):333–58. https://doi.org/10.1504/IJIMA.2012.051615.
3. Bearman M, Tai J, Dawson P, Boud D, Ajjawi R. Developing evaluative judgement for a time of generative artificial intelligence. Assess Evaluation High Educ. 2024;49(6):893–905. https://doi.org/10.1080/02602938.2024.2335321.
4. Brennan J, Naidoo R. Higher education and the achievement (and/or prevention) of equity and social justice. High Educ. 2008;56(3):287–302. https://doi.org/10.1007/s10734-008-9127-3.
5. Castelló-Sirvent F, Roger-Monzó V, Gouveia-Rodrigues R. Quo vadis, university? A roadmap for AI and ethics in higher education. Electron J e-Learning. 2024;22(6):34–51.

https://doi.org/10.34190/ejel.22.6.3267.

6. Colquitt JA. On the dimensionality of organisational justice: A construct validation of a measure. J Appl Psychol. 2001;86(3):386–400. https://doi.org/10.1037/0021-9010.86.3.386.
7. Colquitt JA, Conlon DE, Wesson MJ, Porter COLH, Ng KY. Justice at the millennium: A meta-analytic review of 25 years of organisational justice research. J Appl Psychol. 2001;86(3):425–45. https://doi.org/10.1037/0021-9010.86.3.425.
8. Corbin T, Dawson P, Liu D. Talk is cheap: Why structural assessment changes are needed for a time of GenAI. Assess Evaluation High Educ. 2025a;50(7):1087–97. https://doi.org/10.1080/02602938.2025.2503964.
9. Corbin T, Dawson P, Nicola-Richmond K, Partridge H. Where's the line? It's an absurd line': Towards a framework for acceptable uses of AI in assessment. Assess Evaluation High Educ. 2025b;50(5):705–17. https://doi.org/10.1080/02602938.2025.2456207.
10. DiMaggio PJ, Powell WW. The iron cage revisited: Institutional isomorphism and collective rationality in organisational fields. Am Sociol Rev. 1983;48(2):147–60. https://doi.org/10.2307/2095101.
11. Ellis C, Murdoch K. The educational integrity enforcement pyramid: A new framework for challenging and responding to student cheating. Assess Evaluation High Educ. 2024;49(7):924–34. https://doi.org/10.1080/02602938.2024.2329167.
12. Flutti E, Travaini G, Roma P, Ongis M. Neutralisation techniques and academic dishonesty among Italian university students. Front Psychol. 2026;17., Article 1766612. https://doi.org/10.3389/fpsyg.2026.1766612.
13. Główczewski M, Burdziej S. (2023). (In)justice in academia: Procedural fairness, students' academic identification, and perceived legitimacy of university authorities. *Higher Education*. Advance online publication. https://doi.org/10.1007/s10734-022-00907-8
14. Gonsalves C. Addressing student non-compliance in AI use declarations: Implications for academic integrity and assessment in higher education. Assess Evaluation High Educ. 2024;50(4):592–606. https://doi.org/10.1080/02602938.2024.2415654.
15. Harris PTS. (2024). *Faculty perspectives toward artificial intelligence in higher education* (Publication No. 3102991316) [Doctoral dissertation, Middle Georgia State University]. ProQuest Dissertations & Theses Global. https://www.proquest.com/docview/3102991316
16. Higher Education Policy Institute (HEPI). (2026). *What UK university AI policies actually do: A study of 96 institutions*. https://www.hepi.ac.uk/reports/what-uk-university-ai-policies-actually-do-a-study-of-96-institutions/
17. Jiang Y, Xie L, Cao X. Exploring the effectiveness of institutional policies and regulations for generative AI usage in higher education. High Educ Q. 2025;79. https://doi.org/10.1111/hequ.70054. Article e70054.
18. Leventhal GS. What should be done with equity theory? New approaches to the study of fairness in social relationships. In: Gergen KJ, Greenberg MS, Willis RH, editors. Social exchange: Advances in theory and research. Plenum; 1980. pp. 27–55.

19. Luo J. A critical review of GenAI policies in higher education assessment: A call to reconsider the originality of students' work. Assess Evaluation High Educ. 2024;49(5):651–64. https://doi.org/10.1080/02602938.2024.2309963.
20. Macfarlane B. Teaching with integrity: The ethics of higher education practice. RoutledgeFalmer; 2004.
21. Perkins M, Furze L, Roe J, MacVaugh J. The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. J Univ Teach Learn Pract. 2024;21(1). https://doi.org/10.53761/1.21.1.06.
22. Quality Assurance Agency for Higher Education (QAA). (2023). *Maintaining quality and standards in the ChatGPT era*. https://www.qaa.ac.uk/docs/qaa/members/maintaining-quality-and-standards-in-the-chatgpt-era.pdf
23. Sykes GM, Matza D. Techniques of neutralisation: A theory of delinquency. Am Sociol Rev. 1957;22(6):664–70. https://doi.org/10.2307/2089195.
24. Xia W, Wei W. (2026). Beyond the syllabus: How the fidelity of GenAI governance implementation shapes student trust and transparency behaviours. *Innovations in Education and Teaching International*, Advance online publication. https://doi.org/10.1080/14703297.2026.2655907

Figures

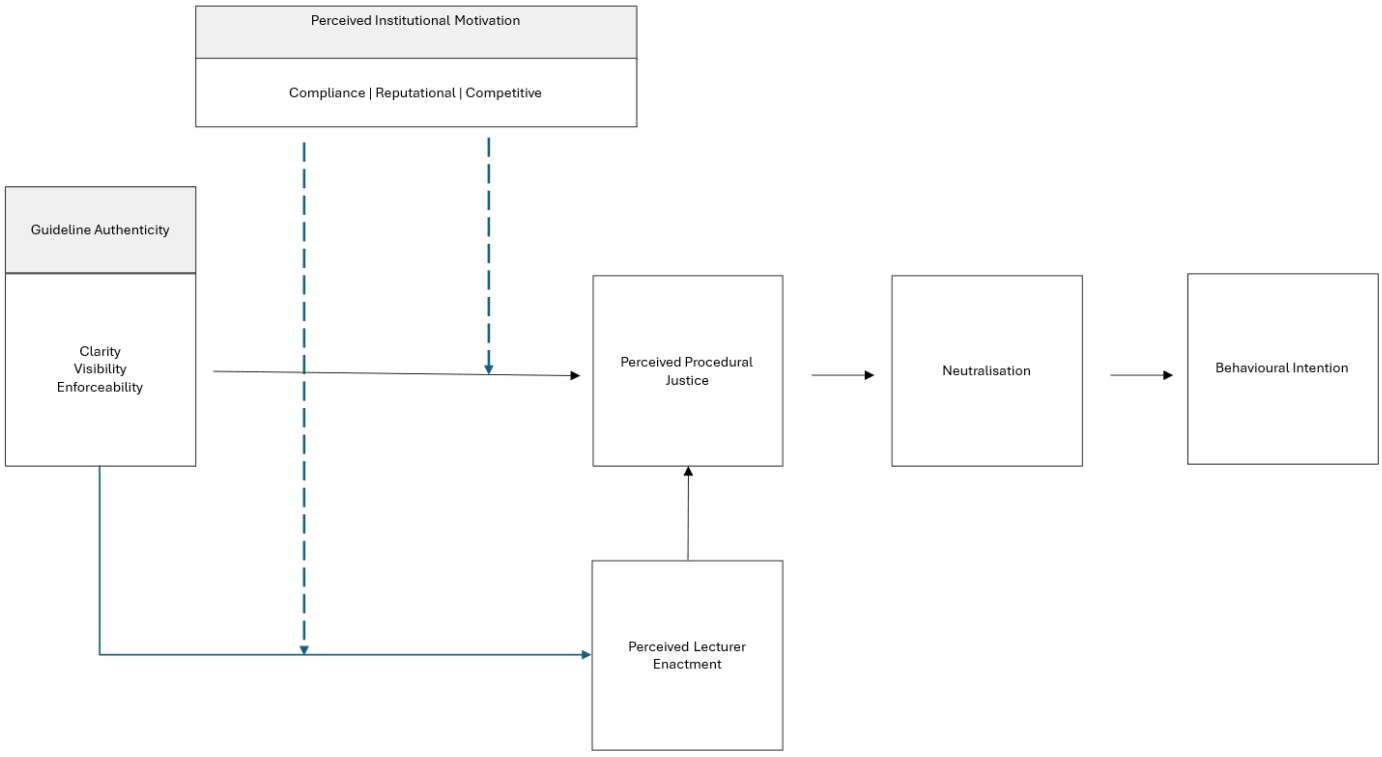


Figure 1

Conceptual Model Showing Direct, Moderation and Mediation Effect Paths

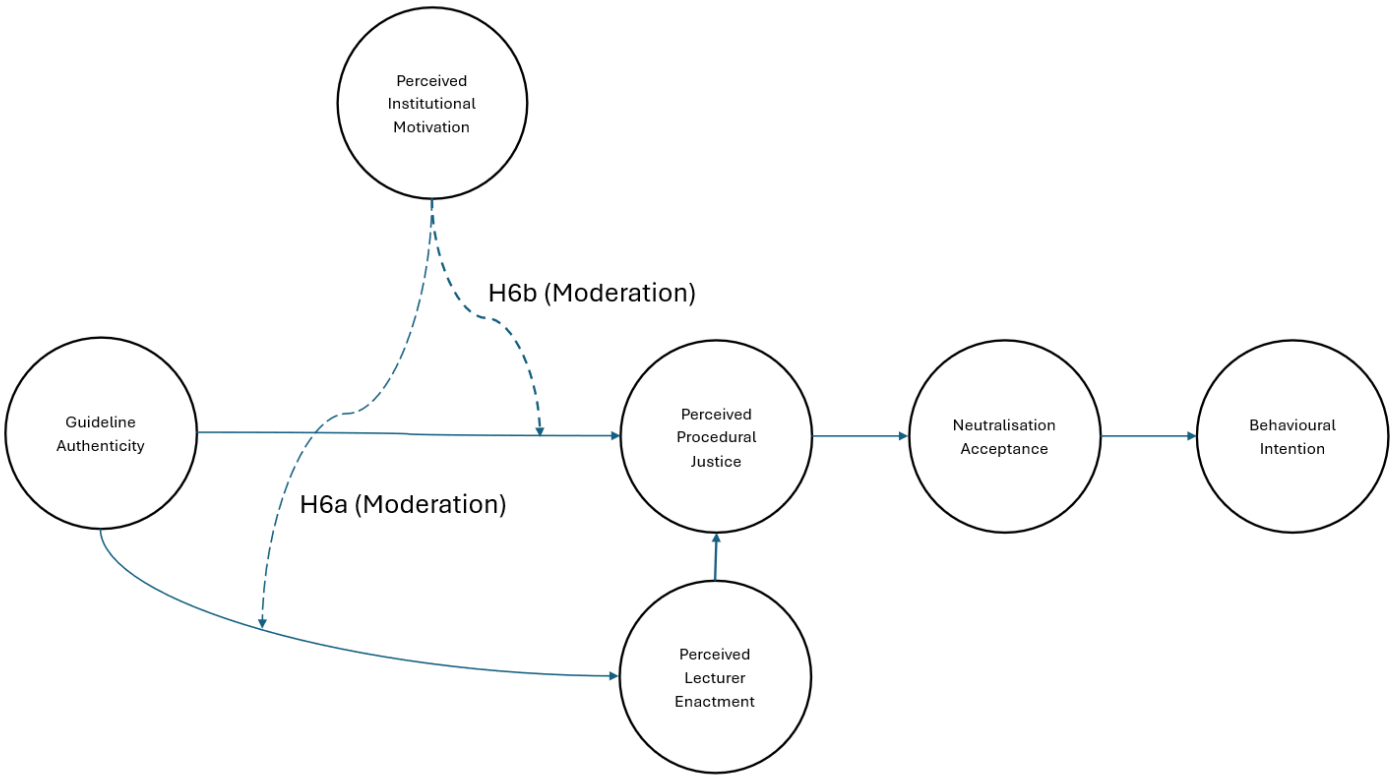


Figure 2

Multi-level Structural Equation Model