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.

The Productivity Myth and Learning Paradox of Generative AI in Higher Education

Muhammad Raheel Anwar

mranwar.mscs22seecs@seecs.edu.pk

National University of Sciences and Technology

Basit Ali

National University of Sciences and Technology

Aqsa Arif

University of Agriculture Faisalabad

Research Article

Keywords: Generative AI, Higher Education, Cognitive Engagement, Assessment Redesign, Academic Integrity, Institutional Policy

Posted Date: June 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9869630/v1

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

Additional Declarations: No competing interests reported.

The Productivity Myth and Learning Paradox of Generative AI in Higher Education

Abstract

Background: The rapid adoption of Generative Artificial Intelligence (GenAI) in higher education has created a structural tension between well-documented productivity benefits and growing concerns about cognitive engagement, academic integrity, and institutional governance. Existing theoretical frameworks have not resolved this tension at scale. This study introduces the Productivity–Cognitive Engagement Paradox to describe the condition in which measurable output gains coexist with critically diminished cognitive involvement and critical thinking.

Methods: A convergent parallel mixed-methods design integrated survey data from 1,037 students and 393 faculty across four disciplines with 40 qualitative interview segments subjected to inductive-deductive thematic analysis (inter-rater $\kappa = 0.84$). Quantitative analysis employed descriptive statistics, multiple regression, Kruskal–Wallis tests, binary logistic regression, and Sobel mediation analysis.

Results: Students reported near-universal high academic productivity ($M = 4.476$) alongside critically low cognitive engagement ($M = 1.988$), a 2.49-point structural divergence. Quadrant analysis placed 83.9% in the Paradox Zone, a pattern robust across all disciplines and usage intensities. No individual-level predictor significantly explained cognitive engagement ($R^2 = .004$, all $p > .22$). Student non-disclosure of AI use was unrelated to academic integrity ($r = .024$, $p = .442$), establishing shadow usage as policy-rational behaviour. Faculty reported near-universal assessment restructuring (89.8% authentic tasks; 82.4% oral examinations), with Institutional AI Policy Clarity as the sole significant predictor of authentic task adoption (OR = 1.816, $p = .023$), yet policy clarity was rated near the scale floor ($M = 2.034$).

Conclusions: Generative AI systematically displaces deep cognitive processes while enhancing academic output. The paradox is structurally robust and individually unmoderated. A tripartite response framework addressing student AI literacy, faculty-led authentic assessment, and institution-level policy governance is proposed as a coordinated, evidence-grounded intervention strategy.

Keywords: Generative AI; Higher Education; Cognitive Engagement; Assessment Redesign; Academic Integrity; Institutional Policy

1 Introduction

The emergence of large language model-based Generative Artificial Intelligence tools has created conditions in higher education that have no clear historical precedent. When OpenAI released ChatGPT in November 2022, universities worldwide confronted a technology capable of producing graduate-level essays, generating functional programming code, synthesising research literature, and completing complex analytical tasks within seconds and at no cost to the student (Adiguzel, Kaya, & Cansu, 2023; Kasneci et al., 2023; OpenAI, 2023). This represented a fundamental discontinuity from prior educational technologies. Whereas earlier tools such as search engines, translation software, or plagiarism detection systems altered how students accessed or verified information, Generative AI altered what students needed to produce independently. The implications of this shift for cognitive development, academic assessment, institutional policy, and the social contract of higher education are profound and remain only partially understood.

Two competing narratives have emerged in the scholarly literature regarding these implications, and the tension between them provides the motivating problem for the present study. The first narrative, well-supported by productivity-focused empirical work, positions Generative AI as a transformative cognitive resource that amplifies student output, personalises learning, and democratises access to expert-level feedback (Lim, Gunasekara, Pallant, Pallant, & Pechenkina, 2023; Mollick & Mollick, 2023). The second narrative, grounded in cognitive load theory and self-determination frameworks, raises the concern that AI tools may produce students who are academically productive but intellectually passive, achieving high-quality outputs without engaging in the deep processing through which genuine learning occurs (Hmelo-Silver, 2004; Ouyang & Jiao, 2021). These two narratives do not merely describe different aspects of the same phenomenon; they potentially describe outcomes that are simultaneously true for the same students.

This simultaneity is what the present study sets out to document and theorise. This study introduces the Productivity–Cognitive Engagement Paradox: an empirically observable condition in which students achieve very high levels of AI-enhanced academic output alongside critically low levels of cognitive engagement, critical thinking, and disciplinary skill development. The paradox is grounded in quantitative data from 1,037 students whose mean productivity score of 4.476 and mean cognitive engagement score of 1.988 on a five-point scale represent a 2.49-point structural divergence that challenges prevailing assumptions about the unambiguous educational benefits of Generative AI adoption.

A third dimension, equally important but less frequently integrated into productivity-and-learning debates, concerns the institutional environment within which students and faculty navigate Generative AI adoption. Faculty are not passively observing these developments; they are restructuring their pedagogical and assessment practices in real time, and doing so largely without coherent institutional guidance (Chan, 2023). This study documents the magnitude of this transformation and examines the consequences of the institutional policy vacuum for both student behaviour and faculty practice.

The study addresses three interrelated knowledge gaps in the existing literature. First, most prior empirical work has examined either student adoption or faculty responses in isolation, whereas the present study examines both simultaneously within a single mixed-methods design, enabling cross-role triangulation of findings. Second, the quantitative literature has disproportionately focused on technology acceptance outcomes (specifically perceived usefulness and ease of use) without examining downstream effects on learning quality, particularly cognitive engagement and critical thinking. Third, the role of institutional AI policy as a systemic moderator of student transparency and faculty efficacy remains empirically understudied at the scale required to draw sector-wide conclusions. By drawing on a large, multi-disciplinary sample with both survey and interview data, this study offers a comprehensive empirical portrait of Generative AI integration in higher education that advances the field on all three fronts.

The remainder of this paper is organised as follows. Section 2 presents the theoretical background and synthesises the relevant literature. Section 3 describes the study context, design rationale, and the four research questions guiding the investigation. Section 4 details the sampling procedures, instruments, and analytical methods. Section 5 presents quantitative and qualitative findings in an integrated sequence. Section 6 interprets the findings in relation to theory and practice. Section 7 identifies limitations and directions for future research, and Section 8 presents conclusions and policy recommendations.

2 Theoretical Background

This section situates the study within four complementary theoretical traditions and reviews empirical evidence across four interconnected domains: Generative AI adoption in higher education, the tension between performance gains and genuine learning, academic integrity and the shadow usage phenomenon, and faculty-led assessment redesign. Together, these strands establish the conceptual and evidential basis for the research design and questions presented in Section 3.

2.1 Theoretical Framework

The conceptual framework guiding this study, illustrated in Figure 1, draws on four complementary theoretical traditions to capture the multi-layered dynamics of Generative AI integration in higher education.

The Technology Acceptance Model, first formulated by Davis (1989), provides the foundation for understanding adoption behaviour. According to this model, individual decisions to adopt a given technology are governed by two primary cognitive evaluations: perceived usefulness and perceived ease of use. In the Generative AI context, these constructs have demonstrated consistent empirical salience: students and faculty who perceive these tools as useful and sufficiently accessible are significantly more likely to adopt and sustain their use (Grassini, 2023; Sallam, 2023). However, the Technology Acceptance Model was designed to explain adoption behaviour, not learning outcomes. The present framework extends the model downstream, treating adoption

as a moderating variable that shapes, rather than determines, the cognitive quality of educational engagement.

Self-Determination Theory, developed by Deci and Ryan (Deci & Ryan, 1985, 2000), argues that deep engagement and intrinsic motivation arise from the satisfaction of three fundamental psychological needs: autonomy, competence, and relatedness. The theory predicts that learning environments supporting autonomous skill-building and progressive competence development will produce richer cognitive engagement than those that bypass independent problem-solving. Generative AI creates a structural tension with this prediction. By resolving cognitive challenges instantly and comprehensively, these tools risk bypassing the competence-building process through which intrinsic engagement is cultivated. When students delegate analytical and compositional tasks to AI systems rather than developing their own capacities, the psychological mechanisms that sustain deep learning are undermined (Ouyang & Jiao, 2021; Ryan & Deci, 2017). This dynamic resonates with Vygotsky (1978)'s concept of the Zone of Proximal Development: learning occurs through effortful engagement at the boundary of current competence, and AI tools risk collapsing that productive boundary by resolving challenges before students engage with them.

Cognitive Load Theory, developed by Sweller (1988), distinguishes between extraneous cognitive load, which burdens working memory without contributing to learning, and germane cognitive load, which represents the productive cognitive effort through which knowledge is structured and encoded into long-term memory. The educational concern raised by Generative AI is not that these tools reduce extraneous load (a potentially beneficial outcome), but that they may also eliminate germane load. The productive struggle through which understanding is built, the iterative drafting, reconceptualising, and critical evaluation that characterise high-quality academic work, may be bypassed when AI generates the output directly. Students who experience this bypass may produce sophisticated work without engaging in the cognitive processes that learning requires (Hmelo-Silver, 2004; Kahneman, 2011).

Institutional Theory, drawing on the foundational work of DiMaggio and Powell (1983), frames the role of organisational norms, regulatory frameworks, and coercive pressures in shaping individual behaviour within formal settings. Applied to Generative AI in higher education, the absence of coherent institutional policies creates a normative vacuum in which neither students nor faculty receive adequate guidance on what constitutes legitimate use. The consequences of such vacuums are well-established in academic integrity research, where ambiguous norms produce behavioural inconsistency, with some individuals engaging in overly restrictive compliance and others defaulting to covert non-compliance (Bretag, 2013; McCabe, Butterfield, & Treviño, 2012). Chan (2023) has explicitly identified institutional policy clarity as a necessary precondition for ethical and productive Generative AI use in educational settings.

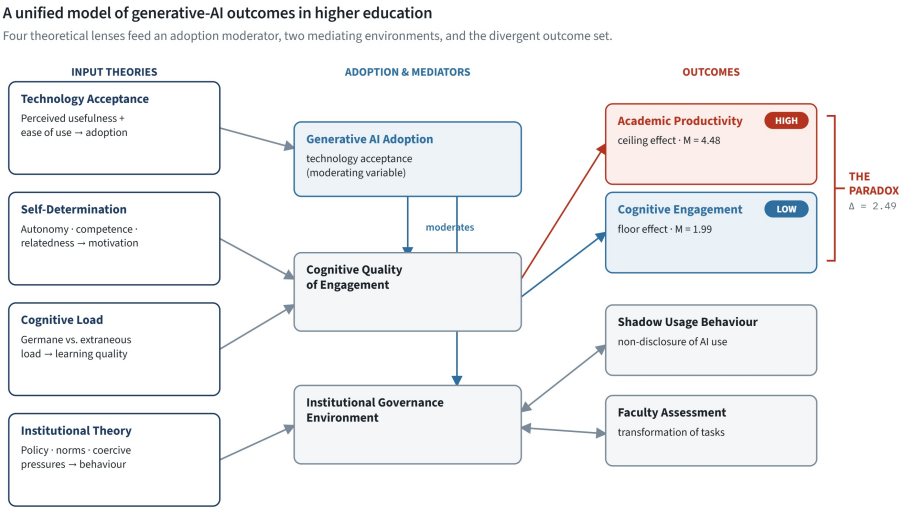


Fig. 1: Conceptual research framework illustrating the integration of Technology Acceptance Model, Self-Determination Theory, Cognitive Load Theory, and Institutional Theory in explaining Generative AI integration outcomes in higher education. Arrows denote proposed causal pathways from technology adoption through cognitive and institutional moderators to educational outcomes.

2.2 Generative AI Adoption in Higher Education

The pace and breadth of Generative AI adoption among university students has exceeded the most optimistic projections made at the time of these tools' public release. Cross-national surveys conducted since 2023 have reported adoption rates ranging from 43% to over 80%, with the most intensive usage concentrated in writing assistance, programming support, and research summarisation (Baidoo-Anu & Owusu Ansah, 2023; Cooper, 2023; Crompton & Burke, 2023; Lo, 2023; Rasul et al., 2023). A systematic review by Zawacki-Richter, Marín, Bond, and Gouverneur (2019) contextualises this wave within a broader pattern of accelerating digital tool adoption that consistently outpaces institutional policy development; a more recent scoping review by Liu, Hu, Gladman, et al. (2025) confirmed this pattern across 39 empirical studies published in the first eight months after ChatGPT's release alone. The dominant tools in this adoption wave (ChatGPT, Microsoft Copilot, Claude, and Google Gemini) differ qualitatively from previous AI-assisted software in that they do not merely support a task but can autonomously complete it. This distinction is theoretically critical, as it shifts the locus of cognitive effort from the student to the system (Bender, Gebru, McMillan-Major, & Shmitchell, 2021).

Faculty adoption of Generative AI has followed a more cautious and ambivalent trajectory. Many instructors report incorporating these tools into administrative workflows (including syllabus preparation, feedback drafting, and resource curation) while maintaining substantive reservations about student use in assessment contexts (Perkins, 2023; Yan et al., 2023). This divergence between student enthusiasm and faculty apprehension is not merely a generational or attitudinal gap; it reflects substantive disagreements about the epistemic integrity of AI-assisted academic work and about the institutional mechanisms adequate to govern it.

2.3 The Performance–Learning Tension

The most consequential and theoretically underexplored dimension of Generative AI in higher education concerns the relationship between the performance it enables and the learning it may preclude. Research on performance outcomes presents a relatively consistent picture. Mollick and Mollick (2023) found that students who used ChatGPT produced essays of measurably higher quality than control groups and completed tasks significantly faster. Lim et al. (2023) documented comparable productivity gains across coding, data analysis, and written synthesis tasks. Farrokhnia, Banihashem, Noroozi, and Wals (2024) identified notable strengths in GenAI's capacity for personalised feedback and adaptive learning support alongside risks of over-reliance and skill attrition. These findings are reproduced with sufficient regularity to be treated as robust, and a recent systematic review of 51 studies confirms that Generative AI exerts large positive effects on learning performance and moderate positive effects on higher-order thinking when used purposively (Hon, 2026).

The learning side of this equation is considerably more complex and less settled. Kasneci et al. (2023) characterised a scaffolding paradox, noting that Generative AI provides support of such power and completeness that students may never develop the independent scaffolding skills that academic tasks were designed to cultivate. Firat (2023) documented that students who habitually rely on AI for essay drafting show a reduced capacity to generate and organise original arguments without technological assistance. Yan et al. (2023) described this process as skill atrophy: the gradual erosion of previously practised capabilities when AI consistently resolves the tasks through which those capabilities are developed and reinforced. Revesai (2025) documented that students exhibiting addictive AI usage patterns showed compulsive engagement alongside continued reliance despite recognising negative academic consequences, while Alhur, Khlaif, Hamamra, and Hussein (2025) found that educators in Palestinian universities independently identified the same paradox: AI dependency undermines the autonomous professional judgment it was intended to support.

These findings, drawn from independent research programmes, converge on the central thesis of the present study. The Productivity–Cognitive Engagement Paradox reflects a structural condition produced by the systematic displacement of learner cognition by AI-generated output, not an artefact of a single study or a narrow sample. Understanding this condition empirically, at sufficient scale to inform policy, is one of the primary contributions of this research.

2.4 Academic Integrity and Shadow Usage

Academic integrity represents one of the most actively debated dimensions of Generative AI adoption in higher education, and the policy responses it has provoked reveal the degree to which existing institutional frameworks are ill-equipped to manage this challenge. University responses since 2023 have ranged from comprehensive prohibition to broadly permissive endorsement, with the majority of institutions occupying an unstable middle ground defined by vague, inconsistently applied guidance (Bittle & El-Gayar, 2025; Perkins, 2023; Tlili, Abed, Bozkurt, et al., 2024; Tlili et al., 2023; UNESCO, 2023).

The consequences of policy ambiguity for student behaviour are well-characterised in the academic integrity literature. Unclear regulations do not produce compliant behaviour; they produce behavioural incoherence, in which students cannot reliably distinguish permissible from impermissible actions and therefore adopt strategies calibrated to risk management rather than academic honesty (Bretag, 2013; McCabe et al., 2012). In the Generative AI context, this dynamic produces what the present study terms *shadow usage*: a pattern of deliberate non-disclosure by students who perceive the regulatory environment as simultaneously ambiguous and punitive. Shadow usage has not yet received dedicated empirical investigation at scale; its measurement in the present study represents a methodological contribution to the field.

2.5 Faculty-Led Assessment Redesign

Faculty responses to the proliferation of Generative AI have been characterised by pragmatic and largely individual innovation in the absence of coordinated institutional direction. The two most widely documented adaptations are the introduction of authentic, project-based assessments requiring personalised knowledge, situated judgment, and iterative development that AI cannot readily replicate, and the reinstatement or expansion of oral examinations as a means of verifying that students can explain, defend, and extend the work they submit (Cotton, Cotton, & Shipway, 2023; Qadir, 2023).

These responses reflect a rational and pedagogically sound diagnosis of the assessment problem. Traditional written assignments, particularly essays or standard coding tasks on predefined topics, can now be completed comprehensively by Generative AI at a quality exceeding many students' independent work. Restructuring assessment to foreground demonstration rather than product, through authentic tasks requiring real-world judgment and oral exchanges demanding spontaneous engagement, restores some of the epistemic validity that AI adoption has eroded.

However, these responses carry substantial implementation challenges that cannot be addressed through individual faculty initiative alone. Oral examinations are resource-intensive and may disadvantage students with communication anxieties, language differences, or disabilities. Authentic task design requires substantial curriculum development investment and may not align with existing accreditation frameworks. Without institutional coordination, students experience radically inconsistent assessment environments across modules and programmes, and faculty lack the professional

development and resource support necessary to implement new approaches at scale (Chan, 2023; Holmes et al., 2022).

3 Research Design

This section establishes the empirical context in which the study was conducted, the rationale for the mixed-methods design, and the four research questions that organised the inquiry. The contextual and design decisions described here directly inform the sampling, instrumentation, and analytical procedures detailed in Section 4.

3.1 Study Context and Design

This study was conducted across multiple higher education institutions in Pakistan during the 2024–2025 academic year, a period characterised by rapidly escalating Generative AI adoption among university students and the emergence of the first institution-level responses to that adoption. Pakistan's higher education sector provides a productive empirical context: it encompasses a diverse mix of public and private universities operating under a central regulatory body (the Higher Education Commission), serves a student population that has embraced Generative AI tools with notable speed and breadth, and exhibits the policy ambiguity regarding AI use that characterises many national systems still in the early stages of governance development.

The multi-institutional design was deliberately constructed to address a fundamental limitation of existing research: studies conducted within a single institution or discipline cannot distinguish between conditions that are structurally universal and those that reflect the specific culture, norms, or pedagogical traditions of a particular setting. Participating institutions spanned four academic disciplines—Social Sciences (25.7%), Business (25.5%), Arts and Humanities (24.7%), and STEM (24.2%)—and included both public (62.5%) and private (37.5%) sector universities. Student participants were recruited across all levels of formal higher education: undergraduate students (69.9%), postgraduate students (24.4%), and doctoral candidates (5.7%). The sample was demographically diverse, spanning four age cohorts—18–20 years (41.1%), 21–23 years (37.4%), 24–26 years (16.6%), and 27 and above (4.9%)—with approximately balanced gender representation: female (48.6%), male (47.3%), and gender-diverse (4.1%). The faculty sample was drawn from the same institutional network, encompassing all major academic ranks—Assistant Professors (26.5%), Associate Professors (25.7%), Lecturers (24.9%), and Professors (22.9%)—and the full range of teaching experience from fewer than five years to more than twenty. This within-study cross-role design enables triangulation of student and faculty perspectives within a shared educational environment, a feature absent from most prior work in this domain.

This study employs a convergent parallel mixed-methods design (Creswell & Plano Clark, 2018), in which quantitative survey data and qualitative interview data are collected independently and simultaneously, then integrated at the stage of interpretation and inference. This design was selected because the research questions

require both the precision of quantifiable measurement and the depth of phenomenological understanding. Constructs such as cognitive engagement and interaction transparency can be operationalised through multi-item Likert-scale instruments, but the mechanisms that produce them—including the motivations underlying shadow usage and the lived experience of AI-induced skill dependency—require qualitative inquiry to be adequately theorised. Triangulating findings across both strands strengthens the internal validity of conclusions and their transferability to comparable institutional contexts. The design is visualised as a conceptual framework in Figure 1.

3.2 Research Questions

The overarching aim of this study is to empirically characterise the structural and institutional dynamics of Generative AI integration in higher education, with specific attention to the tension between AI-enhanced academic output and cognitive engagement, the policy conditions that shape student non-disclosure behaviour, and the factors that drive faculty assessment transformation. The study investigates whether these dynamics constitute a coherent, cross-disciplinary structural pattern or whether they are moderated by individual student characteristics, disciplinary context, or usage behaviour.

Four research questions guide the investigation:

- **RQ1:** What are the patterns of Generative AI adoption among university students; is there empirical evidence of a Productivity–Cognitive Engagement Paradox, and if so, what is its magnitude and distributional character; and do AI dependency, technology acceptance, and self-efficacy predict variation in cognitive engagement when examined together?
- **RQ2:** How have faculty pedagogical and assessment practices transformed in response to Generative AI adoption, and what institutional and individual-level factors drive these transformations?
- **RQ3:** How does institutional AI policy clarity relate to student interaction transparency (shadow usage) and faculty assessment transformation?
- **RQ4:** How do qualitative narratives from students and faculty converge with, and contextualise, the quantitative patterns identified in RQ1–RQ3?

4 Methodology

This section details the participant recruitment and sampling procedures, the survey instruments and interview protocol, and the quantitative and qualitative analytical methods employed to address the research questions.

4.1 Participants and Sampling

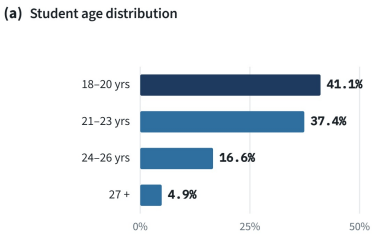
Three participant groups contributed data to this study: undergraduate, postgraduate, and doctoral students; faculty members; and a purposive interview subsample drawn from both groups.

4.1.1 Student participants

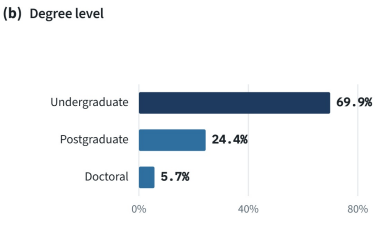
The student sample was recruited through purposive sampling targeting undergraduates, postgraduates, and doctoral candidates enrolled across four academic disciplines at multiple higher education institutions, as described in Section 3. The geographic scope of the sample constitutes a delimiting factor for generalisation; see Section 7 for discussion. Following instrument administration, responses were screened against three quality control criteria: completion of all items, successful response to an embedded attention-check question, and absence of straight-lining (uniform responses across all items within a page block). The retained sample of 1,037 respondents met all three criteria. Full demographic profiles are visualised in Figure 2.

4.1.2 Faculty participants

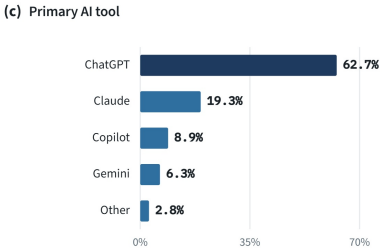
Faculty participants were recruited from institutions aligned with the student sampling frame. After quality checks mirroring those applied to the student data, 393 responses were retained. The faculty sample encompassed all major academic ranks and a full range of teaching experience levels, as detailed in Section 3 and visualised in Figure 2.



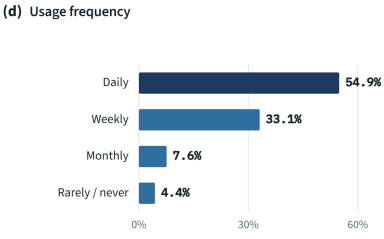
(a) Student age distribution



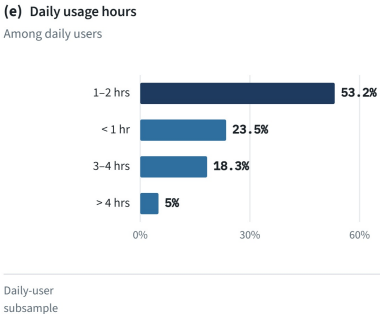
(b) Degree level



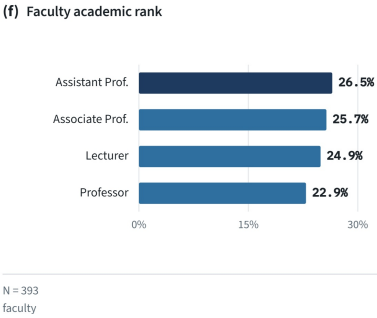
(c) Primary AI tool



(d) Usage frequency



(e) Daily usage hours



(f) Faculty academic rank

Fig. 2: Demographic profiles of student (N = 1,037) and faculty (N = 393) participants. (a) Student age distribution; (b) Degree level composition; (c) Primary AI tool preference; (d) Usage frequency; (e) Daily usage duration; (f) Faculty academic rank distribution.

4.1.3 Interview participants

Qualitative data were collected through semi-structured interviews with a purposive subsample drawn from survey participants who had indicated willingness to engage in further data collection. The qualitative dataset comprises 40 transcribed and thematically coded interview segments, 18 from students and 22 from faculty. Disciplinary and demographic variation was maximised within the interview subsample to ensure that the qualitative findings reflect diverse institutional and professional contexts.

4.2 Instrumentation

Three instruments were developed and piloted for this study: a 63-item student survey, a 36-item faculty survey, and a semi-structured interview protocol. The following subsubsections describe each in turn.

4.2.1 Student survey instrument

The student survey instrument comprised 63 Likert-scale items measuring twelve theoretical constructs on a five-point response format (1 = Strongly Disagree; 5 = Strongly Agree). Perceived Usefulness was assessed using four items adapted from Davis (1989), and Perceived Ease of Use with a corresponding four-item scale from the same source. AI Literacy was operationalised through a five-item scale adapted from Ng, Leung, Chu, and Qiao (2021). Information Quality External (IQE) was assessed using five items capturing students' perceptions of the reliability and accuracy of AI-generated content relative to conventional academic sources; this construct serves as a contextual complement to the dependency scale. AI Dependency (AID) was captured with four items developed originally for this study; items were labelled IQI (Information Quality–Internal) in the raw instrument but are referred to throughout this paper as AID to more precisely reflect the construct's theoretical content as functional reliance on AI to initiate and complete cognitive tasks. Cognitive Engagement was assessed using four items adapted from Fredricks, Blumenfeld, and Paris (2004), and Self-Efficacy with four items derived from Bandura (1986). Critical Thinking was captured through five items adapted from Facione (1990). Learning Effectiveness and Academic Productivity were measured by five and three items respectively, both developed for this study. Academic Integrity was assessed using five items adapted from McCabe et al. (2012), and Interaction Transparency was operationalised through three items developed to capture patterns of self-disclosure regarding AI use. Demographic and AI usage items completed the instrument.

A pilot study administered to 40 participants was conducted prior to full rollout. Participant feedback confirmed adequate item clarity across the majority of constructs, and minor revisions were made to the Interaction Transparency subscale in response to phrasing queries.

4.2.2 Faculty survey instrument

The faculty instrument comprised 36 items measuring six constructs. Observed Student Impact on Performance was assessed using five items. Faculty Trust in standard

student submissions was measured through six items adapted from Mayer, Davis, and Schoorman (1995). Pedagogical Transformation and Assessment Redesign were measured with three and five items respectively. Institutional AI Policy Clarity was assessed using four items adapted from Chan (2023), and Faculty Institutional Influence through three items. Binary response items documented whether each faculty member had adopted oral examinations and authentic task assessments.

4.2.3 Semi-structured interview protocol

The interview protocol comprised twelve open-ended questions organised around three thematic areas: personal experience with Generative AI tools, observations of AI's effects on student learning and assessment quality, and institutional factors shaping AI-related practice. All interviews were audio-recorded with informed consent, transcribed verbatim, and uploaded to NVivo 14 for systematic coding.

4.3 Analytical Procedures

Quantitative analysis. Construct scores were computed as unweighted means of their constituent indicator items. Descriptive statistics were calculated for all constructs, including means, standard deviations, medians, and skewness values. Independent-samples t-tests examined differences between daily and non-daily AI users and between STEM and non-STEM students on key outcome constructs. Cohen's d was computed to assess practical effect sizes. Multiple regression analyses identified significant predictors of Cognitive Engagement and Academic Productivity. All analyses were performed in Python 3.12 (Python Software Foundation, 2023) using the pandas (McKinney, 2010), scipy (Virtanen, Gommers, Oliphant, et al., 2020), and numpy (Harris, Millman, van der Walt, et al., 2020) libraries.

Qualitative analysis. The 40 interview segments were analysed using a hybrid inductive-deductive approach following the six-phase thematic analysis procedure described by Braun and Clarke (2006). A preliminary codebook was developed deductively from the theoretical framework, then progressively refined as emergent themes from the data extended or complicated the initial structure. Each segment received codes for a primary theme, a secondary theme where applicable, sentiment orientation, and a linked research hypothesis. Thematic saturation was assessed iteratively: no substantively new codes emerged after the 32nd segment, and the final eight segments produced only confirmatory instances of existing codes, providing adequate evidential basis for the six themes reported. To establish inter-rater reliability, a secondary coder with doctoral-level qualifications in educational research independently coded 25% of the corpus (ten randomly selected segments). Agreement between raters was assessed using Cohen's kappa, which yielded $\kappa = 0.84$, indicating strong inter-rater concordance.

Integration strategy. Quantitative and qualitative findings were integrated using a convergence matrix adapted from Creswell and Plano Clark (2018). Instances in which both strands confirmed the same conclusion were treated as triangulated evidence. Instances in which qualitative data extended or complicated the quantitative patterns were explicitly theorised to generate novel interpretive propositions.

4.4 Ethical Considerations

Ethical approval for this study was obtained from the relevant Institutional Review Board prior to data collection. All survey and interview participants were provided with a detailed participant information sheet explaining the study's purpose, data handling procedures, and their right to withdraw at any time without consequence. Written informed consent was obtained from all participants before data collection commenced. All data were anonymised at the point of collection; no personally identifying information was retained in the analysis datasets. Interview recordings and transcripts were stored in a password-protected repository accessible only to the research team. Participation in both the survey and interview components was entirely voluntary, and no incentives that could be considered coercive were offered. The study was conducted in full accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments.

5 Results

This section presents findings from both methodological strands in an integrated sequence. The quantitative results are organised by research question: RQ1 findings span AI adoption patterns, paradox characterisation, and cognitive engagement modelling; RQ2 and RQ3 address faculty assessment transformation and the institutional policy vacuum; cross-group universality analyses support all three. The qualitative thematic analysis (RQ4) follows, triangulating and extending the quantitative patterns.

5.1 Quantitative Findings

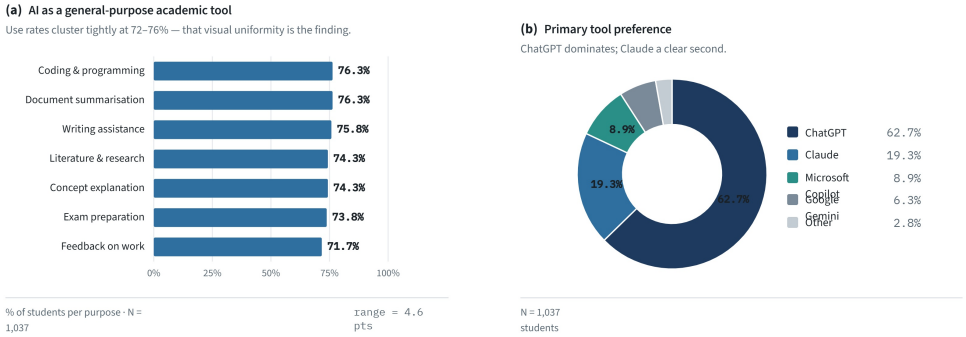
Quantitative results are presented in sequence. Findings addressing RQ1 are reported first, spanning AI adoption patterns, paradox characterisation and distributional analysis, regression modelling of cognitive engagement, and cross-group universality tests. Faculty assessment transformation (RQ2) and the institutional policy vacuum (RQ3) follow. The qualitative thematic analysis (RQ4) concludes the section.

5.1.1 Generative AI adoption patterns (RQ1)

Generative AI adoption among the student sample was both extensive and intensive. ChatGPT was identified as the primary tool by 62.7% of students, followed by Claude (19.3%), Microsoft Copilot (8.9%), and Google Gemini (6.3%). A majority reported daily use (54.9%), with 33.1% using AI tools weekly and 7.6% monthly. Among daily users, the modal usage duration was one to two hours per day (53.2%), while 18.3% reported three to four hours and 5.0% exceeded four hours of daily engagement.

Generative AI was employed across a remarkably wide range of academic purposes with strikingly uniform usage rates across categories: writing assistance (75.8%), coding and programming support (76.3%), literature and research facilitation (74.3%), document summarisation (76.3%), concept explanation (74.3%), feedback on submitted work (71.7%), and examination preparation (73.8%). This near-uniform distribution (spanning a narrow range of 71.7%–76.3%) indicates that Generative AI

has been integrated as a general-purpose academic tool rather than one associated with particular task types, a finding with substantial implications for discipline-specific assessment design. The full adoption profile is presented in Figure 3.



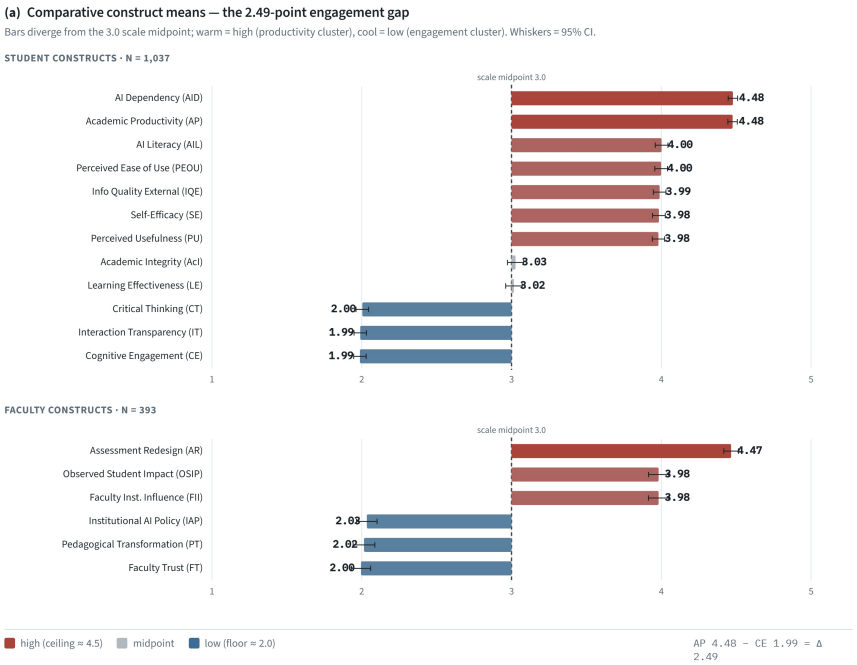
(a) AI usage purpose rates across seven academic task categories

(b) Primary AI tool preference distribution across disciplines

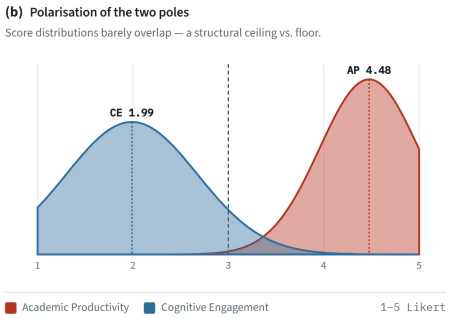
Fig. 3: Generative AI adoption profile among student participants (N = 1,037). (a) Usage purpose rates across seven academic task categories, showing near-uniform 72–76% adoption across all purposes; (b) Primary tool preference distribution, with ChatGPT dominating at 62.7% followed by Claude (19.3%), Copilot (8.9%), and Gemini (6.3%).

5.1.2 The Productivity–Cognitive Engagement Paradox (RQ1)

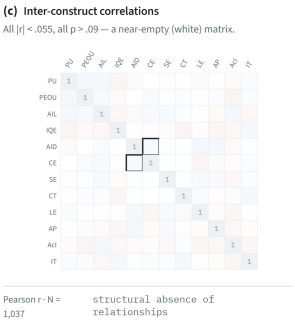
The central empirical finding of this study is the dramatic and theoretically significant polarisation in student outcome construct scores. Descriptive statistics for all twelve student constructs are presented in Table 1 and visualised comparatively in Figure 4.



(a) Comparative construct means for all student and faculty scales, illustrating the 2.49-point divergence between Academic Productivity ($M = 4.476$) and Cognitive Engagement ($M = 1.988$)



(b) Paradox polarisation: distribution of productivity and engagement scores, highlighting the near-universal ceiling and floor effects



(c) Inter-construct correlation matrix for student variables, showing the near-zero relationship between AI Dependency and cognitive outcomes

Fig. 4: The Productivity–Cognitive Engagement Paradox and supporting analyses. (a) Construct mean comparison across all student ($N = 1{,}037$) and faculty ($N = 393$) variables with 95% confidence intervals; (b) Score distribution plot illustrating structural polarisation; (c) Inter-construct correlation matrix.

Table 1: Descriptive statistics for student constructs ($N = 1{,}037$)

Construct	M	SD	Min	Max	Median	Skewness
Perceived Usefulness (PU)	3.981	0.674	2.00	5.00	4.00	−0.24
Perceived Ease of Use (PEOU)	3.999	0.694	1.50	5.00	4.00	−0.29
AI Literacy (AIL)	4.002	0.696	1.00	5.00	4.00	−0.29
Info Quality External (IQE)[c]	3.989	0.691	1.80	5.00	4.00	−0.28
AI Dependency (AID)[a]	4.478	0.519	3.00	5.00	4.50	−0.47
Cognitive Engagement (CE)	1.988	0.686	1.00	4.50	2.00	+0.24
Self-Efficacy (SE)	3.984	0.698	2.00	5.00	4.00	−0.34
Critical Thinking (CT)	2.004	0.694	1.00	4.00	2.00	+0.25
Learning Effectiveness (LE)	3.016	0.901	1.00	5.00	3.00	+0.01
Academic Productivity (AP)	4.476	0.519	3.00	5.00	4.67	−0.41
Academic Integrity (AcI)[b]	3.027	0.885	1.00	5.00	3.00	+0.05
Interaction Transparency (IT)	1.990	0.682	1.00	4.00	2.00	+0.29

Note. All items measured on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Cronbach's α ranged from .916 (Academic Productivity) to .972 (Learning Effectiveness), indicating excellent internal consistency across all constructs. [a]AI Dependency (AID) was operationalised through items originally labelled IQI (Information Quality–Internal) in the instrument; AID is used throughout this paper to more accurately reflect the construct's theoretical content. [b]The Academic Integrity construct is abbreviated AcI in correlation figures to avoid confusion with the acronym AI (Artificial Intelligence). [c]Information Quality External (IQE) is included for descriptive completeness; it was not selected as a focal predictor in the regression analyses given the study's primary emphasis on dependency, engagement, and transparency constructs.

Academic Productivity registered the highest student construct mean at $M = 4.476$ ($SD = 0.519$), with a negative skew of −0.41. This distribution indicates that elevated productivity is not a central tendency characterising an average student but a near-universal condition across the sample, with almost no student reporting productivity below the scale midpoint. This finding converges with the established productivity literature (Lim et al., 2023; Mollick & Mollick, 2023) and confirms that the performance benefits of Generative AI are broad-based rather than restricted to technologically proficient or highly motivated subgroups.

Cognitive Engagement, by contrast, registered a mean of $M = 1.988$ ($SD = 0.686$), nearly 2.5 scale points below Academic Productivity and substantially below the theoretical midpoint of 3.0. Critical Thinking scores were comparably low at $M = 2.004$ ($SD = 0.694$). Taken together, these scores indicate that the capacity for reflective, effortful, and self-directed cognitive processing has been substantially displaced across the sample as a whole. The 2.49-point gap between Academic Productivity and Cognitive Engagement constitutes the empirical core of the Productivity–Cognitive Engagement Paradox.

The AI Dependency construct yielded the joint-highest score in the student dataset at $M = 4.478$ ($SD = 0.519$), reflecting a near-ceiling distribution: students comprehensively report an inability or unwillingness to initiate academic tasks without AI assistance. The Interaction Transparency construct produced the joint-lowest mean in the dataset at $M = 1.990$ ($SD = 0.682$), indicating that systematic non-disclosure of AI use was the modal, not exceptional, pattern. The qualitative analysis in Section 5.2 provides phenomenological grounding for this shadow usage pattern as rational risk-management under conditions of policy ambiguity.

Information Quality External (IQE), measuring students' perceptions of AI-generated content as reliable relative to conventional academic sources, yielded a moderate mean of $M = 3.989$ ($SD = 0.691$), above the scale midpoint but substantially lower than both AI Dependency and AI Literacy. This profile suggests that students broadly accept AI-generated content as adequate for academic purposes, though their functional dependency ($M = 4.478$) outstrips their confidence in AI output quality, indicating that usage is not straightforwardly attributable to uncritical trust in AI reliability.

To quantify the paradox's prevalence, students were classified into four quadrants based on whether their Academic Productivity and Cognitive Engagement scores fell above or below the scale midpoint of 3.0 (see Figure 5). The results are striking: 870 of 1,037 students (83.9%) occupied the Paradox Zone (AP ≥ 3, CE < 3), a combination of simultaneously high productivity and low engagement. Only 164 students (15.8%) achieved the theoretically ideal condition of both high productivity and high engagement (AP ≥ 3, CE ≥ 3). Very few students exhibited engaged-but-unproductive or uniformly low profiles. The large concentration in the Paradox Zone (83.9% of a 1,037-student sample) establishes the Productivity–Cognitive Engagement Paradox as a dominant structural condition. Bivariate analysis found a near-zero association between AI Dependency and Cognitive Engagement: $r(\text{AID}, \text{CE}) = +.037$, $p = .231$, consistent with the near-universal character of disengagement regardless of dependency levels.

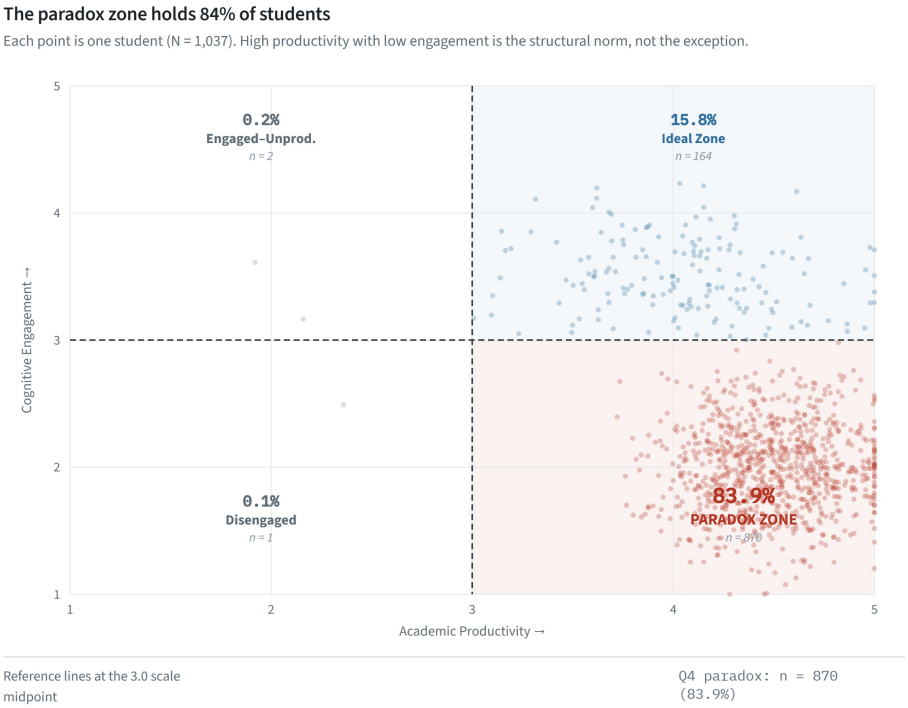


Fig. 5: Paradox Zone quadrant scatter plot (N = 1,037). Students are classified by Academic Productivity (x-axis) and Cognitive Engagement (y-axis) relative to the scale midpoint of 3.0. Q4 (Paradox Zone: high productivity, low engagement) contains 870 students (83.9%). Q1 (Ideal Zone: high on both) contains 164 students (15.8%). Q2 and Q3 together account for a small minority (0.3%).

5.1.3 Predictors of Cognitive Engagement (RQ1)

Multiple regression analysis was conducted with Cognitive Engagement as the dependent variable and Perceived Usefulness, AI Literacy, AI Dependency, Self-Efficacy, and Learning Effectiveness as predictors. Results are presented in Table 2 and visualised in Figure 6.

Note. $R^2 = .004$, Adjusted $R^2 = -.001$, $F(5, 1031) = 0.83$, $p = .530$. No predictor reached statistical significance (all $p > .22$). B = unstandardised coefficient; β = standardised coefficient.

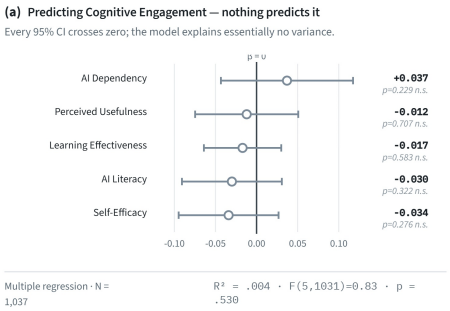
No predictor reached statistical significance in the regression on Cognitive Engagement (all $p > .22$; overall model: $F(5, 1031) = 0.83$, $p = .530$, $R^2 = .004$). This result is theoretically meaningful rather than merely null. With a sample of 1,037 participants, the test had substantial statistical power to detect even small effects; the absence of significance thus indicates that individual-level variation in technology acceptance, AI literacy, dependency, or self-efficacy does not systematically explain variance in cognitive engagement. This is coherent with the paradox's structural character: when 83.9% of students are in the Paradox Zone regardless of their individual profiles, no

Table 2: Multiple regression analysis: Cognitive Engagement as dependent variable ($N = 1{,}037$)

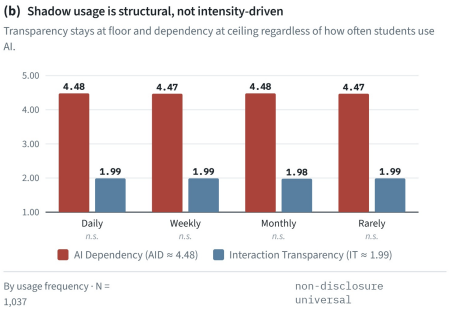
Predictor	B	SE	β	t	p
Intercept	2.107	0.287	—	7.346	$< .001$
Perceived Usefulness	−0.012	0.032	−.012	−0.375	.707
AI Literacy	−0.030	0.031	−.030	−0.992	.322
AI Dependency (AID)	+0.049	0.041	+.037	+1.203	.229
Self-Efficacy	−0.033	0.031	−.034	−1.090	.276
Learning Effectiveness	−0.013	0.024	−.017	−0.549	.583

single predictor can meaningfully differentiate disengaged from engaged students. The non-significant coefficient for AI Dependency ($B = +0.049$, $p = .229$) specifically indicates that dependency level, while near-universally high, does not further differentiate engagement within the sample.

A parallel regression analysis examining predictors of Academic Productivity found no statistically significant predictors among the technology acceptance variables (all $p > .37$, $R^2 = .001$). This null finding is consistent with the hypothesis that productivity gains are a near-universal baseline effect of AI capability that accrues broadly across users regardless of individual attitudes, literacy levels, or self-efficacy. The full AP regression output is available from the corresponding author upon request.



(a) Regression coefficients for predictors of Cognitive Engagement ($N = 1{,}037$). No predictor is statistically significant (all $p > .22$), consistent with the structural interpretation that CE variation is not explained by individual-level differences in the sample.



(b) Interaction Transparency and AI Dependency scores by usage frequency group, illustrating the shadow usage pattern

Fig. 6: Regression and shadow usage patterns. (a) Standardised regression coefficients for predictors of Cognitive Engagement ($N = 1{,}037$); no predictor is significant, consistent with CE being a near-universal structural condition; (b) Interaction Transparency items ($M = 1.990$), all scoring below 2.0, confirming that shadow usage is a structural condition rather than an intensive-user artefact.

5.1.4 Mediation analysis

To examine whether AI Dependency mediated the pathway from daily usage frequency to Cognitive Engagement (the hypothesised mechanism through which intensive use produces cognitive disengagement), a Sobel mediation analysis was conducted with AI Dependency as the proposed mediating variable. Path a (daily usage frequency to AI Dependency) was non-significant ($a = -0.005$, $p = .786$). Path b (AI Dependency to Cognitive Engagement) was equally non-significant ($b = +0.049$, $p = .231$). The Sobel test confirmed the absence of a statistically significant indirect effect ($z = -0.265$, $p = .791$). The path diagram is visualised in Figure 7. This null mediation result is theoretically coherent with the structural interpretation of the paradox: the near-ceiling distribution of AI Dependency ($M = 4.478$, $SD = 0.519$) compresses individual-level variance to the point where pathway detection via conventional mediation methods is statistically implausible. Mediation analysis cannot reveal a functional pathway through a variable that has effectively saturated. The finding reinforces the conclusion that the Productivity–Cognitive Engagement Paradox is a population-level structural condition rather than an individually modulated phenomenon.

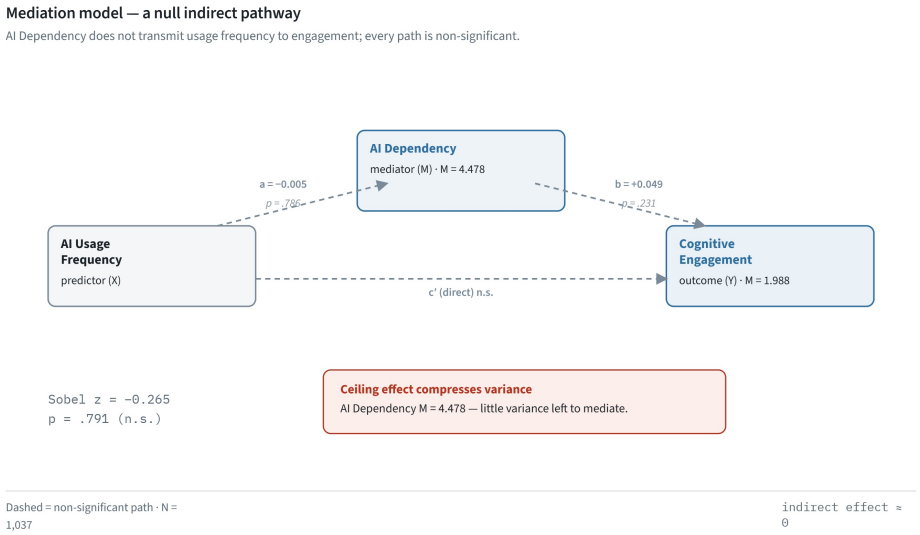


Fig. 7: Path diagram for the mediation model (AI Usage Frequency → AI Dependency → Cognitive Engagement). Neither path reached significance: the path from usage frequency to AI Dependency ($a = -0.005$, $p = .786$) and the path from AI Dependency to Cognitive Engagement ($b = +0.049$, $p = .231$) were both non-significant, with Sobel $z = -0.265$, $p = .791$. The ceiling effect on AI Dependency ($M = 4.478$) compresses variance and limits the pathway's detectability.

5.1.5 Assessment Transformation and the Policy Vacuum

Descriptive statistics for all six faculty constructs are presented in Table 3, with the assessment transformation profile visualised in Figure 9.

Table 3: Descriptive statistics for faculty constructs ($N = 393$)

Construct	M	SD	Min	Max	Median	Skewness
Observed Student Impact (OSIP)	3.982	0.680	2.00	5.00	4.00	−0.21
Faculty Trust in Submissions (FT)	1.995	0.652	1.00	4.00	2.00	+0.11
Pedagogical Transformation (PT)	2.015	0.736	1.00	4.00	2.00	+0.32
Assessment Redesign (AR)	4.466	0.493	3.00	5.00	4.40	−0.27
Institutional AI Policy (IAP)	2.034	0.694	1.00	4.00	2.00	+0.35
Faculty Institutional Influence (FII)	3.982	0.689	2.00	5.00	4.00	−0.32

Note. All items measured on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Cronbach's α ranged from .928 (Faculty Institutional Influence) to .964 (Observed Student Impact), indicating excellent internal consistency across all faculty constructs.

Faculty Trust in standard student submissions was critically low at $M = 1.995$ ($SD = 0.652$), the joint lowest faculty construct score alongside Pedagogical Transformation. This near-floor level of institutional confidence in the integrity of submitted academic work indicates a systemic epistemological crisis in the assessment process rather than merely a generalised attitudinal shift. Assessment Redesign, at $M = 4.466$ ($SD = 0.493$), represents the inverse response to this trust erosion. High levels of assessment redesign activity characterise virtually the entire faculty sample, not merely a reformist minority. These two constructs in combination (critically low trust paired with intensively high redesign activity) map onto a clear institutional dynamic in which faculty have diagnosed a fundamental failure in traditional assessment validity and are responding with comprehensive restructuring.

The binary adoption data confirms the scale of this transformation. Authentic task assessments had been adopted by 89.8% of faculty, and oral examinations had been incorporated by 82.4%. These figures represent a near-universal restructuring of assessment methodology within the sample. Institutional AI Policy Clarity yielded a mean of $M = 2.034$ ($SD = 0.694$), similarly low to Faculty Trust. The institutional apparatus that should be coordinating and governing this transformation is substantially absent, with faculty engaging in largely uncoordinated pedagogical innovation without the professional development, resource allocation, or policy scaffolding that would render these innovations coherent and equitable.

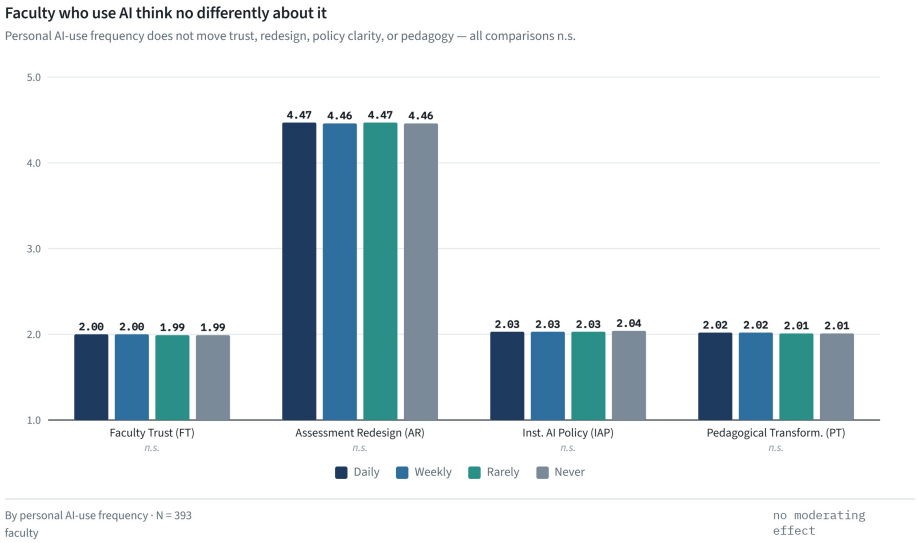
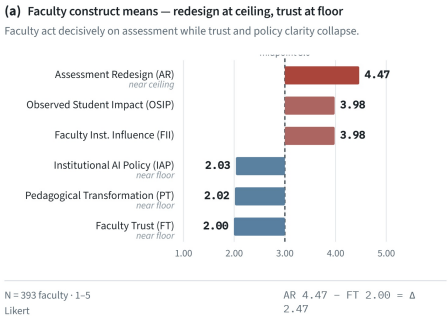


Fig. 8: Faculty construct scores (Faculty Trust, Assessment Redesign, Institutional AI Policy, Pedagogical Transformation) stratified by personal AI use frequency (Daily, Weekly, Rarely, Never). No statistically significant moderation of faculty constructs by personal AI use frequency was identified, indicating that faculty who use AI themselves do not differ substantially in their institutional trust or assessment redesign behaviour from those who do not.

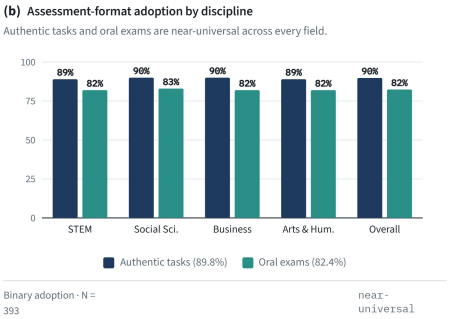
A theoretically significant asymmetry is observable between the Pedagogical Transformation ($M = 2.015$, $SD = 0.736$) and Assessment Redesign ($M = 4.466$, $SD = 0.493$) constructs. Despite the near-universal adoption of restructured assessment formats, faculty report critically low levels of holistic pedagogical transformation. This asymmetry suggests that the faculty response is primarily defensive and assessment-driven, a targeted fix to a validity problem, rather than a broader reconceptualisation of pedagogical philosophy. Faculty are redesigning how learning is evaluated without similarly rethinking how it is facilitated. The Observed Student Impact (OSIP, $M = 3.982$) and Faculty Institutional Influence (FII, $M = 3.982$) constructs both occupy the moderate-to-high range, indicating that faculty perceive meaningful effects of AI on student performance and retain a degree of institutional voice, an important contextual observation given that both Trust and Policy Clarity are near the floor.

Binary logistic regression was conducted to identify predictors of authentic task adoption (Table 4). For authentic task adoption, Institutional AI Policy Clarity was the sole statistically significant predictor ($B = 0.596$, OR $= 1.816$, 95% CI [1.08, 3.04], $p = .023$), indicating that faculty at institutions with clearer AI policy frameworks were more likely to adopt authentic assessment formats. Academic rank, Faculty Trust, personal AI use, and years of teaching experience were all non-significant predictors. Pseudo-R^2 (McFadden) $= .041$. A parallel model for oral examination adoption did not identify any statistically significant predictors, indicating that oral examination

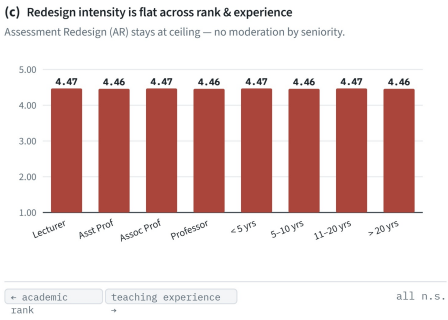
uptake (already near-universal at 82.4%) was not further differentiated by institutional or individual faculty-level variables in this sample; full results are available from the corresponding author. This finding positions institutional governance as a direct enabler of assessment transformation: clarifying the policy environment not only influences student behaviour but also facilitates faculty innovation.



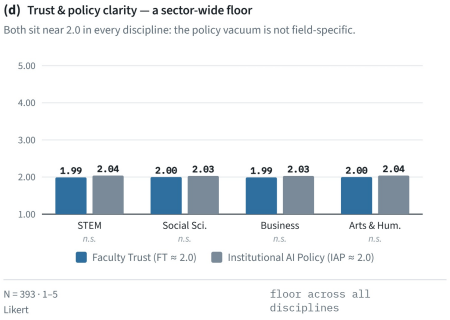
(a) Faculty construct means: trust, policy, redesign, and transformation scores



(b) Adoption rates for authentic task assessments and oral examinations by discipline



(c) Faculty assessment redesign intensity by academic rank and teaching experience



(d) Faculty Trust in student submissions versus Institutional AI Policy Clarity by discipline

Fig. 9: Faculty assessment transformation and the institutional crisis. (a) Faculty construct means illustrating the trust–redesign polarity; (b) Discipline-level adoption rates for alternative assessment formats; (c) Assessment redesign intensity disaggregated by academic rank and experience; (d) Faculty Trust and Institutional Policy Clarity by discipline, confirming the sector-wide policy vacuum.

Note. OR = odds ratio. Bold indicates $p < .05$. Pseudo-R^2 (McFadden) = .041.

Faculty personal AI use frequency did not significantly moderate any faculty construct; construct-level scores were comparable across daily, weekly, infrequent, and

Table 4: Binary logistic regression: predictors of authentic task assessment adoption ($N = 393$)

Predictor	B	OR	95% CI	p
Institutional Policy (IAP)	0.596	1.816	[1.08, 3.04]	**.023**
Academic Rank	0.178	1.195	[0.52, 2.73]	.673
Faculty Trust (FT)	−0.272	0.762	[0.46, 1.27]	.298
Assessment Redesign (AR)	0.363	1.438	[0.74, 2.79]	.283
Teaching Experience	0.245	1.278	[0.94, 1.74]	.116
Personal AI Use	−0.142	0.868	[0.60, 1.27]	.461

non-using faculty (Figure 8), confirming that the assessment transformation and policy vacuum documented here characterise the faculty population broadly rather than being driven by faculty AI adopters specifically.

5.1.6 Correlations and Shadow Usage

The full Pearson correlation matrix among student constructs (Figure 10) reveals a striking pattern of near-zero bivariate associations. All correlations among constructs are statistically non-significant (all $|r| < .055$, all $p > .09$), including the relationship between AI Dependency and Cognitive Engagement ($r = +.037$, $p = .231$). This absence of significant relationships across the construct matrix is theoretically coherent with the structural paradox: when both AI Dependency and Cognitive Engagement show near-ceiling and near-floor effects respectively, cross-individual variation is compressed and conventional regression relationships cannot emerge. The correlation between Interaction Transparency and Academic Integrity was $r = +.024$ ($p = .442$), effectively zero. This null finding is theoretically significant: students who are least transparent about their AI use are not, by that quality, more likely to violate academic integrity norms. Shadow usage, systematic non-disclosure of AI involvement, appears to be a policy-rational behaviour driven by ambiguous institutional regulation rather than an expression of underlying integrity deficit. This reframing has direct implications for how institutions should design their response to the shadow usage phenomenon, as elaborated in Section 6.2.

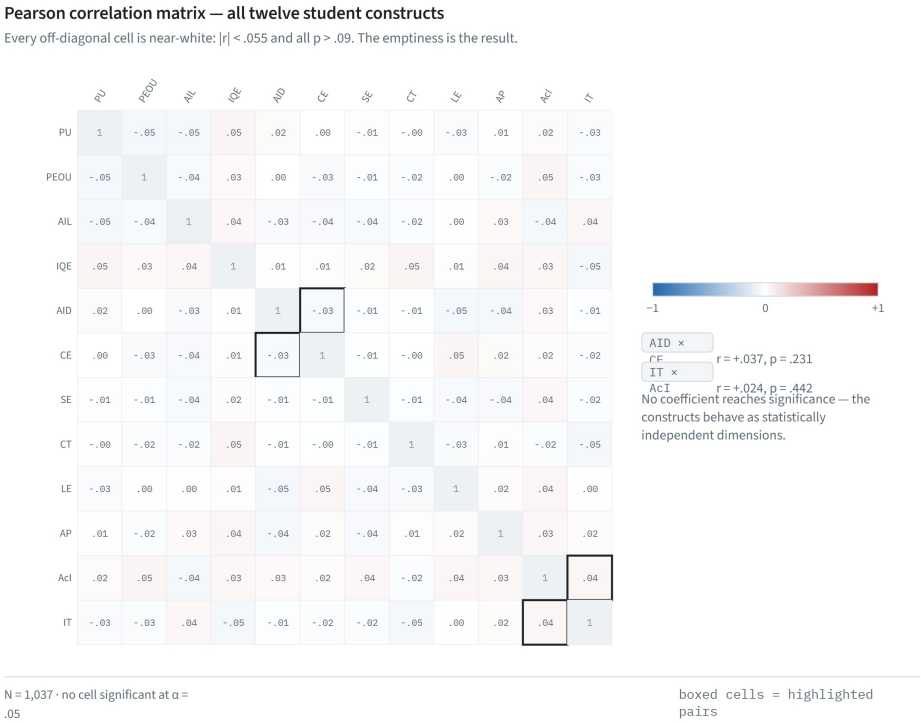


Fig. 10: Pearson correlation heatmap for all twelve student constructs ($N = 1{,}037$). Colour intensity represents correlation magnitude. No correlation is statistically significant (all $|r| < .055$, all $p > .09$). The near-zero $r(\text{AID, CE}) = +.037$ ($p = .231$) and $r(\text{IT, AcI}) = +.024$ ($p = .442$) (where AcI denotes Academic Integrity, abbreviated to avoid confusion with the acronym AI) confirm that disengagement and shadow usage are structural rather than individually driven phenomena.

5.1.7 Cross-group universality

To assess whether the Productivity–Cognitive Engagement Paradox was moderated by student characteristics, Kruskal–Wallis H tests were conducted across three grouping variables: academic discipline (four categories: STEM, Social Sciences, Arts and Humanities, Business), cumulative GPA band (four categories), and self-reported daily AI usage hours (four categories). Results are summarised in Table 5.

Note. ε^2 = epsilon-squared effect size. Discipline differences are non-significant on all constructs (all $p > .11$). CGPA band differences are significant for Academic Productivity and AI Dependency; Cognitive Engagement remains non-significant across CGPA bands, confirming that disengagement is uniform irrespective of academic achievement. AI Dependency differs by daily usage hours; CE does not. $^{*}p < .05$.

Table 5: Kruskal–Wallis tests: cognitive engagement, academic productivity, and AI dependency by student characteristics

Grouping Variable	Construct	H	df	p	ε^2
Discipline (4 groups)	Cognitive Engagement	4.13	3	.248	.001
	Academic Productivity	2.62	3	.454	.000
	AI Dependency (AID)	2.28	3	.516	.000
	Critical Thinking	5.89	3	.117	.003
CGPA Band (4 groups)	Cognitive Engagement	0.05	3	.997	.000
	Academic Productivity	11.24	3	.011*	.008
	AI Dependency (AID)	8.50	3	.037*	.005
Daily Hours (4 groups)	Cognitive Engagement	0.56	3	.906	.000
	AI Dependency (AID)	10.15	3	.017*	.007
	Critical Thinking	1.22	3	.749	.000

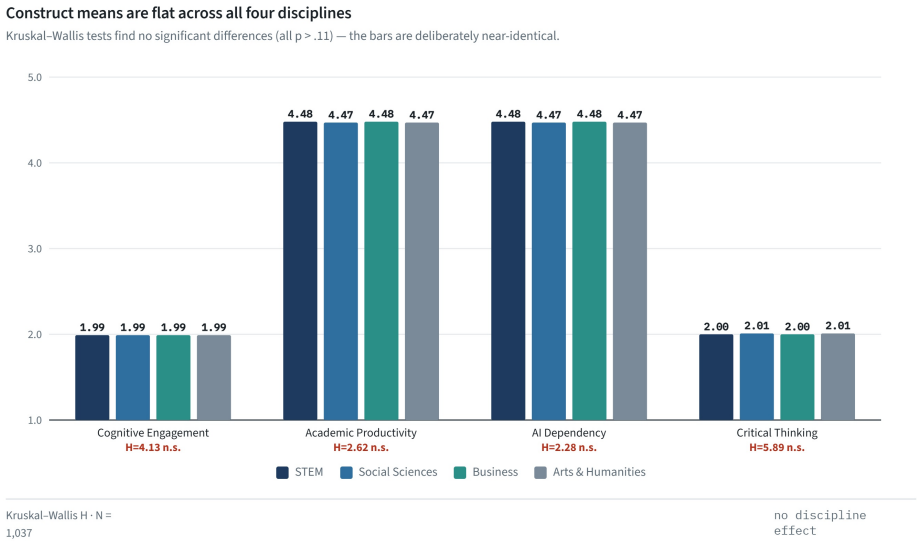


Fig. 11: Kruskal–Wallis comparisons of Cognitive Engagement and Academic Productivity across four academic disciplines (STEM, Social Sciences, Arts and Humanities, Business). No statistically significant discipline differences were found on any construct (all $p > .11$), confirming the cross-disciplinary universality of cognitive disengagement. Error bars represent 95% confidence intervals.

Cognitive Engagement showed no statistically significant differences across any grouping variable (discipline, CGPA band, or usage hours; all $p > .24$), confirming that cognitive disengagement is structurally universal regardless of student characteristics. Academic Productivity did differ significantly across CGPA bands ($H = 11.24$, $p =$

.011, $\varepsilon^2 = .008$), with higher-CGPA students reporting slightly lower productivity means (CGPA 3.6–4.0: $M = 4.46$) than lower-CGPA peers (CGPA < 2.5: $M = 4.63$), suggesting that high-achieving students may apply AI tools more selectively. AI Dependency similarly differed across CGPA bands ($H = 8.50$, $p = .037$) and showed a dose-response relationship with daily usage hours ($H = 10.15$, $p = .017$, $\varepsilon^2 = .007$), with dependency increasing from $M = 4.46$ for users reporting fewer than one hour daily to $M = 4.68$ for those reporting more than four hours. Critically, CE remained near-constant across all usage-hour categories ($M \approx 2.00$), indicating that no usage threshold exists below which engagement is preserved. These findings are visualised in Figures 11, 12, and 13.

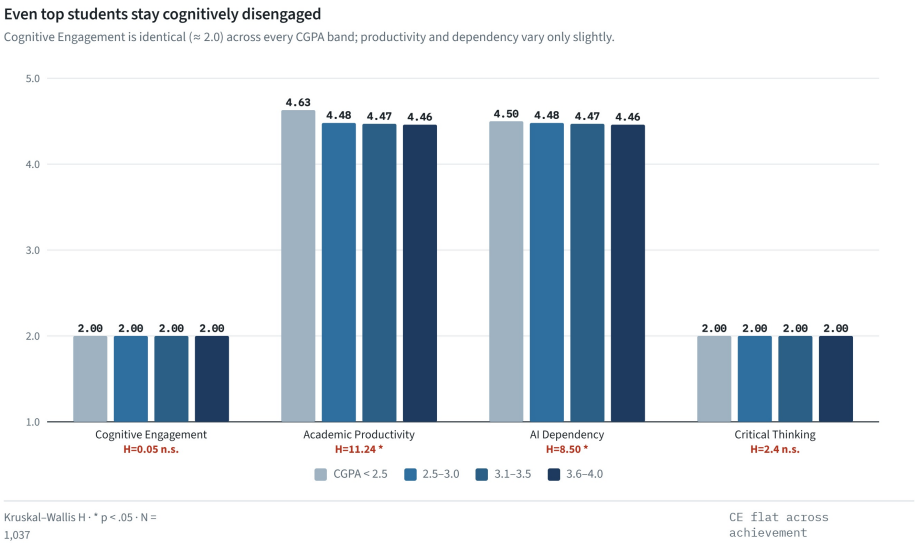


Fig. 12: Cognitive Engagement, Academic Productivity, AI Dependency, and Critical Thinking scores disaggregated by CGPA band. Kruskal–Wallis tests found no significant difference in Cognitive Engagement across CGPA groups ($p = .997$), establishing that high-achieving students are as cognitively disengaged as lower-achieving peers. Academic Productivity ($p = .011$) and AI Dependency ($p = .037$) showed statistically significant but small variation across CGPA groups.

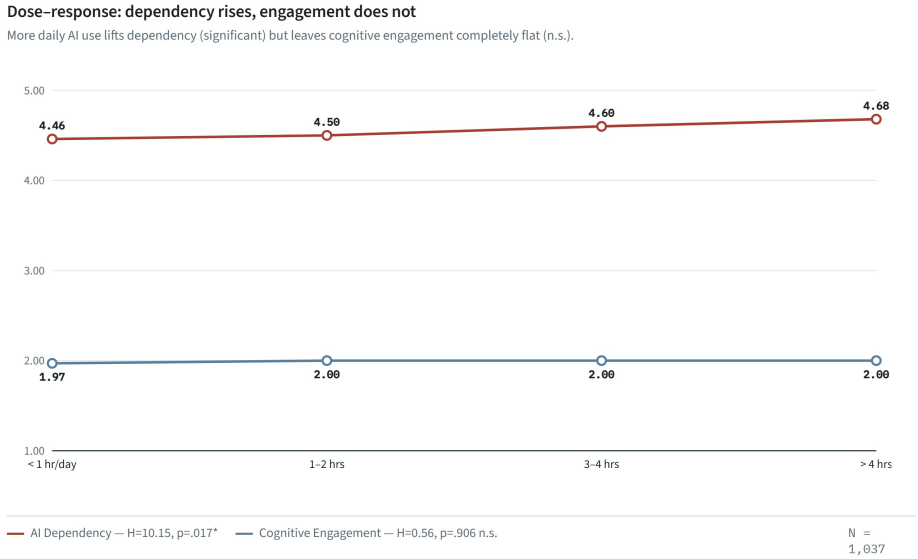


Fig. 13: Dose–response analysis of Cognitive Engagement and AI Dependency across self-reported daily AI usage categories (< 1 hr, 1–2 hrs, 3–4 hrs, >4 hrs). Cognitive Engagement showed no significant dose-response relationship ($H = 0.56$, $p = .906$); even students reporting < 1 hour of daily use registered $M = 1.97$. AI Dependency increased significantly with usage hours ($H = 10.15$, $p = .017$, $\varepsilon^2 = .007$), rising from $M = 4.46$ to $M = 4.68$ across the usage range.

5.1.8 Group comparisons

An independent-samples t-test compared daily AI users ($n = 569$) with less-frequent users ($n = 468$) on key outcome constructs. Daily users reported marginally higher cognitive engagement ($M = 2.02$ vs. $M = 1.95$; $t = 1.708$, $p = .088$, Cohen's $d = 0.107$) and marginally higher academic productivity ($M = 4.49$ vs. $M = 4.46$; $t = 0.736$, $p = .462$, $d = 0.046$). Neither difference reached statistical significance, and effect sizes were negligible to small. AI Dependency scores were effectively identical across groups ($M = 4.47$ vs. $M = 4.48$; $t = -0.226$, $p = .821$, $d = -0.014$), confirming that near-ceiling dependency characterises the full student sample rather than only the most intensive users. The absence of meaningful group differences is consistent with the Kruskal–Wallis results: frequency of use does not systematically moderate cognitive engagement, productivity, or dependency at the group level.

5.2 Qualitative Findings (RQ4)

Thematic analysis of the 40 interview segments, contributed by 18 students and 22 faculty, yielded six major themes. Sentiment distribution across the qualitative corpus showed positive orientations in 40.0% of segments, negative orientations in 32.5%, and neutral orientations in 27.5%. This more balanced distribution, compared with

the productive-concerns polarity common in the literature, reflects a genuine ambivalence in which substantive appreciation for AI's educational benefits coexists with professional and academic concerns about its systemic effects. The thematic network is visualised in Figure 14.

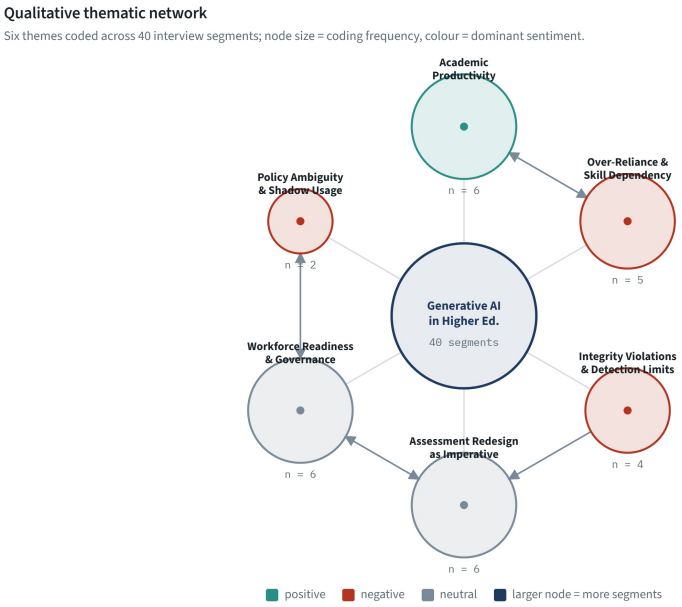


Fig. 14: Qualitative thematic network derived from 40 interview segments. Node size represents coding frequency; arrows indicate thematic relationships. The central node (Generative AI in Higher Education) connects six peripheral themes: Academic Productivity, Over-Reliance and Skill Dependency, Policy Ambiguity and Shadow Usage, Assessment Redesign, Academic Integrity Violations, and Workforce Readiness and Institutional Governance.

Theme 1: Academic productivity as a transformative benefit ($n = 6$ segments; predominantly student; positive). Students across disciplines described AI-enabled productivity gains in concrete and experientially specific terms that went beyond generalised enthusiasm. One student in a STEM discipline (S11, postgraduate) noted that AI tools allow the summarisation of lengthy research documents in seconds, freeing time for the critical analysis that the summarisation was intended to facilitate. Another (S03, Social Sciences, undergraduate) described having access to a responsive explanatory resource functioning outside scheduled instruction hours, providing personalised clarification of complex conceptual material on demand. These accounts converge with the quantitative Academic Productivity mean of 4.476 and lend it phenomenological substance.

A particularly notable observation arising from this theme is that students frame productivity gains as unambiguously positive without spontaneously acknowledging any associated cognitive trade-off. No student in this theme group identified the potential erosion of independent reasoning capacity as a concern requiring attention. This metacognitive gap (the absence of perceived disengagement among students who are, by the quantitative evidence, substantially disengaged) carries important implications for intervention design. Students cannot self-regulate a trade-off they do not perceive; effective intervention must therefore be structurally embedded in curriculum design rather than dependent on individual student reflection.

Theme 2: Over-reliance and skill dependency ($n = 5$ segments; student; negative). Several students described awareness of functional dependency even while contextualising it within normalised usage patterns. One student in a Business discipline (S17, undergraduate) reflected that heavy reliance on AI-assisted coding had diminished confidence in the ability to produce even basic programming constructs independently. This account illustrates with precision the skill atrophy dynamic documented by Firat (2023) and Yan et al. (2023): the withdrawal of AI support reveals an absence of underlying competence that the student had not noticed developing.

This theme provides crucial qualitative grounding for the near-universal AI Dependency mean ($M = 4.478$, approaching the scale ceiling) and the paradox zone distribution. Although the regression found no statistically significant effect of AID on CE in the full sample, the qualitative evidence confirms that individual students experience the dependency–disengagement dynamic in ways that are phenomenologically real and personally consequential. The absence of a regression effect is coherent with the structural interpretation: when both AID and CE are near-ceiling and near-floor respectively across the whole sample, individual variation cannot be expected to reveal a cross-individual relationship. Importantly, even in this theme, students do not frame dependency as a problem requiring intervention but as a pragmatic characteristic of a tool-rich learning environment, suggesting that the metacognitive gap identified in Theme 1 extends even to students who have partial awareness of skill erosion.

Theme 3: Policy ambiguity and shadow usage ($n = 2$ segments; student; negative). Although this theme is supported by fewer qualitative segments than others, reflecting the greater reluctance of students to speak openly about non-disclosure behaviour, the accounts that emerged were substantively rich and internally consistent with the quantitative pattern. The shadow usage pattern documented quantitatively in the Interaction Transparency construct ($M = 1.990$) receives direct phenomenological explanation in this theme. Students across disciplines and institution types described a consistent decision process in which the perceived risk of disclosure outweighed any perceived benefit. One student (S08, Arts and Humanities, undergraduate) noted that the vagueness of institutional policies created conditions in which even relatively minor AI assistance (such as grammar correction) could not be confidently distinguished from behaviour that might be penalised. This account instantiates precisely the theoretical prediction derived from institutional theory, where ambiguous normative frameworks do not produce cautious compliance but instead generate strategically evasive behaviour calibrated to minimise personal risk (Bretag, 2013; DiMaggio & Powell, 1983).

The consistency of this narrative across participants from different disciplines and institutions strengthens the inference that shadow usage is not an individual ethical failure but a rational systemic response to inadequate institutional governance. This reframing has substantial implications for how universities should design their AI policy interventions, as detailed in Section 6.2.

Theme 4: Assessment redesign as pedagogical imperative ($n = 6$ segments; predominantly faculty; neutral). Faculty across disciplines characterised assessment redesign not as an experimental choice but as a professional necessity imposed by the practical inadequacy of existing assessment formats. One faculty member (F14, Associate Professor, Social Sciences) stated that the traditional written essay had ceased to function as a valid instrument for evaluating student understanding in a context where AI systems can produce essay-quality text on demand. The migration to oral validation and in-class presentations was described not as a methodological preference but as a response to a fundamental collapse in the epistemic validity of document-based assessment.

The characterisation of traditional written assessment as no longer viable is a striking claim with substantial disciplinary and sector-wide implications. If oral and authentic assessments are genuinely necessary conditions for reliable evaluation in an AI-rich educational environment, the assessment systems of most universities require comprehensive redesign rather than incremental adjustment. The quantitative Assessment Redesign mean of 4.466 and binary adoption rates of 89.8% and 82.4% suggest this redesign is already substantially underway at the faculty level. The question raised by this theme is whether institutional structures are adequate to support, coordinate, and govern what is, in practical terms, a wholesale transformation of academic evaluation.

Theme 5: Academic integrity violations and detection limitations ($n = 4$ segments; faculty; negative). Faculty who had encountered direct evidence of AI-generated academic submissions described experiences combining professional concern with a sense of institutional inadequacy. One faculty member (F06, Senior Lecturer, STEM) recounted identifying AI-generated programming code by questioning students on their submissions and finding that they were unable to explain the logical structure of the functions they had ostensibly authored. The oral examination, implicitly invoked in this account, emerges from qualitative evidence not merely as one assessment format among others but as the evidentiary instrument through which AI-assisted submission is identified.

A second faculty participant (F19, Professor, Business) framed the detection challenge in terms of a structural asymmetry between generative and detection technologies: the iterative improvement cycles of Generative AI tools consistently outpace those of detection software, rendering detection-based integrity strategies systematically unreliable. This characterisation reflects a genuine and widely recognised technological reality. The qualitative evidence from this theme provides experiential grounding for the statistical pattern of near-universal assessment redesign: faculty have shifted their practices not because detection has failed in isolated cases but because they regard the entire detection-based paradigm as epistemically untenable.

Theme 6: Workforce readiness and the imperative for institutional governance (n = 6 segments; mixed; neutral). A distinct and moderating perspective emerged across both student and faculty participants, foregrounding the professional development implications of Generative AI competence rather than its risks to academic integrity. One faculty participant (F22, Associate Professor, Business) articulated the view that prohibitionist institutional responses to AI are professionally counterproductive, given that graduates are entering workplaces in which fluent AI use is increasingly an expected and valued professional capability. This perspective positions the policy challenge not as a choice between permitting and restricting AI use, but as the institutional responsibility to develop structured frameworks within which AI is used as a professional skill rather than a covert academic shortcut.

The workforce readiness perspective, though addressed less frequently in the existing empirical literature than the preceding assessment and integrity themes, carries proportionally significant weight in the policy implications of this study. Alhur et al. (2025) independently identified this dual imperative in a qualitative study of medical educators, finding that faculty simultaneously recognised AI dependency as professionally problematic and its competent use as professionally necessary, precisely the tension that defines the governance challenge. It suggests that the resolution of the Productivity–Cognitive Engagement Paradox is not an argument against Generative AI in education but an argument for its thoughtful, scaffolded, and institutionally governed integration.

6 Discussion

The following discussion integrates findings from both methodological strands to address the study's core theoretical and policy questions. The paradox's mechanism and structural implications are examined first, followed by the consequences of the institutional policy vacuum, the divergence between assessment transformation and broader pedagogical change, and the scale and equity challenges that assessment transformation must navigate. Cross-group universality findings then establish the population-level scope of these challenges before a tripartite framework for coordinated institutional response is proposed in Section 6.6.

6.1 Interpreting the Paradox

The 2.49-point structural divergence between Academic Productivity and Cognitive Engagement documented in the Results cannot be adequately explained by any single prior theoretical framework. The Technology Acceptance Model accounts for why students adopt and value these tools but was not designed to model the downstream consequences of adoption for cognitive quality. Cognitive Load Theory accounts for the mechanism — the displacement of germane cognitive load when AI generates output directly — but cannot independently predict the magnitude of this effect or the conditions under which it is most pronounced. Self-Determination Theory accounts for the motivational consequences of bypassing competence-building processes but does not specify which tool features or usage patterns trigger the bypass. The integrated framework employed in this study, which situates technology acceptance within

a self-determination and cognitive load account of learning consequences, provides explanatory reach that no individual theory can supply alone.

The near-universal null regression result — no individual-level predictor of Cognitive Engagement reaching significance in a well-powered sample of over 1,000 — is theoretically meaningful rather than merely inconclusive. When the overwhelming majority of students occupy the same structural quadrant regardless of their individual profiles, the expected architecture of a regression relationship cannot emerge because there is no meaningful within-sample variation in the outcome to explain. The paradox is better understood as a population-level floor effect than as a variable that individuals can be positioned above or below through different attitudes or behaviours. This framing has a direct implication for intervention design: individual-targeted approaches calibrated to identify and support cognitively at-risk students are structurally misspecified. The condition is not concentrated in an at-risk subgroup; it pervades the population.

The qualitative evidence deepens this interpretation by revealing the paradox's metacognitive dimension. Students in Theme 1 describe productivity gains enthusiastically without perceiving any corresponding erosion of independent capability. Students in Theme 2 describe skill dependency with concern but without connecting it to the broader pattern of cognitive displacement, or having any framework to address it. The paradox is, in this sense, phenomenologically invisible to those experiencing it. This observation reinforces the structural interpretation: if students could perceive their disengagement and self-regulate accordingly, we would expect individual-level variation to appear. The uniform disengagement across the sample, combined with the qualitative evidence of widespread metacognitive gaps, suggests that the condition is structural and self-sustaining. Self-regulation strategies are therefore insufficient. Students cannot regulate their way out of a structural condition they do not perceive. Effective intervention must be embedded in curriculum design, assessment architecture, and pedagogical scaffolding (Ng et al., 2021; Ouyang & Jiao, 2021).

6.2 The Institutional Policy Vacuum

The simultaneous observation that Institutional AI Policy Clarity and Faculty Trust are both near the floor of their respective scales, while Assessment Redesign and AI Dependency are at the ceiling, reveals a systemic failure in higher education's institutional response to Generative AI that is structural rather than incidental. The institutional policy vacuum is not a temporary administrative lag that will resolve itself as awareness grows; it is an active condition that generates and sustains the very behaviours it ostensibly seeks to prevent.

Institutional theory predicts precisely this outcome. In the absence of clearly articulated norms, individuals do not default to the most ethical or most prudent available behaviour; they default to risk management strategies calibrated to the specific institutional incentives and penalties they can perceive (DiMaggio & Powell, 1983). For students operating in an environment where AI use policies are vague, enforcement is inconsistent, and the costs of disclosure are uncertain, the rational strategy is non-disclosure. Systematic non-disclosure of AI use is not an expression of dishonesty as

a character trait but of rational responsiveness to an ambiguous regulatory environment. A policy response that frames shadow usage as an integrity failure and responds punitively misdiagnoses the problem: the diagnosis is institutional, and the remedy is clear, positively framed policy guidance. A systematic review by Bittle and El-Gayar (2025) of 41 GenAI-integrity studies reached precisely this conclusion: institutional governance frameworks that define rather than merely prohibit AI use are the most effective lever for aligning student behaviour with academic integrity norms.

For faculty, the policy vacuum produces pedagogical isolation and professional overextension. The near-universal assessment restructuring documented in the Results is occurring without the institutional coordination that would render it coherent, equitable, and sustainable. The logistic regression result that Institutional AI Policy Clarity is the significant predictor of authentic task adoption (OR = 1.816, $p = .023$) directly links governance quality to assessment innovation, positioning policy clarity not merely as a student-behaviour moderator but as an active enabler of faculty-led educational reform. Without such clarity, assessment expectations diverge radically across modules and programmes, and the student experience is governed by the initiative of individual educators rather than a shared institutional rationale.

6.3 Assessment Redesign Without Pedagogical Transformation

The striking divergence between Pedagogical Transformation ($M = 2.015$) and Assessment Redesign ($M = 4.466$) is one of the most theoretically revealing findings in the faculty data. Faculty are comprehensively restructuring assessment, near-universally adopting oral examinations and authentic tasks, while reporting no broader transformation in their pedagogical philosophy or practice. This asymmetry suggests that the current faculty response is essentially defensive: a targeted and rational repair of a specific validity failure rather than a reconceptualisation of how learning should be facilitated in an AI-rich environment. Assessment has been redesigned to close the loophole that AI has opened, but the pedagogy that would help students develop genuine competencies has not been correspondingly transformed.

This finding qualifies the optimism that might otherwise accompany the near-universal assessment adoption rates. The transformation is real and necessary, but it is downstream of the cognitive engagement problem rather than upstream of it. Redesigned assessment can detect the absence of learning; it cannot by itself cultivate the presence of it. A fuller institutional response to the Productivity–Cognitive Engagement Paradox requires the pedagogical transformation the data shows has not yet occurred, including active learning designs, scaffolded AI-integration exercises, and metacognitive development programmes that address the root condition rather than its assessment symptoms.

6.4 Scale, Equity, and Institutional Coordination

The scale of the assessment transformation documented in this study is remarkable. Within a two-to-three year window following the broad public release of large language models, nearly nine in ten faculty respondents had fundamentally restructured their

approach to student evaluation. This represents not incremental pedagogical adjustment but structural transformation at a pace that higher education systems have seldom experienced in so short a timeframe (Liu et al., 2025; Zawacki-Richter et al., 2019). The transformation is driven by a clear and rational diagnosis: if AI systems can produce the kinds of outputs that traditional assessments were designed to elicit, those assessments no longer measure what they were designed to measure and must be replaced by formats that restore epistemic validity.

However, this assessment transformation as currently constituted carries two classes of risk that institutional coordination must address. The first concerns equity and accessibility. Oral examinations may disadvantage students whose academic abilities are genuine but whose communicative fluency in the language of instruction is limited, or whose anxiety in evaluative speaking contexts undermines performance that is otherwise competent. Authentic task assessments, while pedagogically richer, may disadvantage students from lower socioeconomic backgrounds who have less access to the real-world contexts, professional networks, and material resources that such tasks sometimes presuppose. These equity concerns are not arguments against restructuring assessment; they are arguments for doing so with institutional care, explicit equity frameworks, and appropriate student support structures.

The second risk concerns coherence and transferability. When assessment redesign is conducted entirely at the individual faculty level, students encounter a patchwork of practices without coherent principles, shared expectations, or institutional rationale. The considerable educational value of authentic assessment and oral examination is partially offset when students cannot understand why they are being evaluated differently across courses or what standard of performance each format is designed to assess. Institutional coordination through shared frameworks for assessment design, faculty professional development, and transparent communication with students about the rationale for new approaches is necessary to realise the full pedagogical potential of the transformation already underway.

6.5 Policy as the Primary Reform Driver

The need for a coordinated institutional response is sharpened by the cross-group universality of the paradox's structural characteristics. Kruskal–Wallis testing found that Cognitive Engagement was non-significantly different across all three grouping variables examined — discipline, CGPA band, and daily usage hours — with effect sizes indistinguishable from zero. The statistically significant but substantively trivial variation in AI Dependency across CGPA and usage categories ($\varepsilon^2 < .01$) does not constitute moderation of the paradox itself; the CE construct that defines the core problem remains flat across all groupings. The intervention implication is unambiguous: targeted strategies designed for at-risk subgroups (lower-achieving students, heavy users, specific disciplines) are structurally misspecified because cognitive disengagement does not concentrate in identifiable subpopulations. It characterises the population of Generative AI users in higher education as a whole.

The near-zero correlation between Interaction Transparency and Academic Integrity forecloses the intuitive interpretation that shadow usage is a proxy for integrity-compromised behaviour. Consistent with institutional theory (Bretag, 2013;

DiMaggio & Powell, 1983), non-disclosure is a rational response to an ambiguous regulatory environment, not an expression of character-level dishonesty. Institutions that respond to shadow usage primarily through punitive enforcement are therefore treating a policy-generated symptom as if it were character-driven — a category error that will produce compliance anxiety without resolving the underlying governance failure.

On the faculty side, the logistic regression finding that Institutional AI Policy Clarity is the significant predictor of authentic task adoption (OR = 1.816, $p = .023$), rather than any individual faculty characteristic, places institutional governance at the centre of the reform dynamic. The analysis establishes that what drives faculty to implement authentic assessments is not personal autonomy or seniority but the institutional scaffolding that makes structured innovation feasible. This has direct policy implications: rather than waiting for reform to diffuse through senior faculty ranks, institutions can accelerate the transformation by developing clear, actionable AI policy frameworks.

6.6 A Tripartite Response Framework

The convergent evidence from quantitative and qualitative analyses points toward a response framework structured around three interdependent intervention strands. These strands are not sequential stages in a reform process but simultaneously necessary components of a coherent institutional response to a systemic challenge.

The first strand concerns *student AI literacy and metacognitive development*. The metacognitive gap identified throughout this study, namely the widespread inability of students to perceive the cognitive trade-offs they are making, indicates that AI literacy education must extend beyond tool operation and prompt engineering to encompass critical reflection on how AI use shapes cognition. Educational programmes should explicitly address the distinction between AI-assisted performance and independent cognitive engagement, incorporating pedagogical activities such as structured self-monitoring assessments and deliberate practice exercises without AI assistance that make this distinction experientially concrete (Ng et al., 2021; Ouyang & Jiao, 2021).

The second strand concerns *faculty-led authentic assessment design and institutional coordination*. The largely individual assessment transformation documented in this study should be formalised, resourced, and coordinated through institutional teaching and learning units. Shared frameworks for authentic assessment design would provide faculty with evidence-based principles and practical tools, rather than requiring each educator to develop new approaches in isolation. Communities of practice structured around assessment innovation would enable faculty to exchange experience and develop shared norms that can be communicated clearly to students (Biggs, 2003; Cotton et al., 2023; Dai et al., 2024). Formal recognition of assessment redesign as a scholarly activity would create the institutional incentives necessary to sustain innovation over time.

The third strand concerns *institution-level AI policy governance*. Clear, transparent, and positively framed AI policies (specifying the conditions under which AI assistance is appropriate, defining the boundaries between permissible assistance and misrepresentation, and communicating these boundaries consistently across all institutional channels) are a structural prerequisite for ethical and productive AI use. Such

policies should be developed collaboratively with students, faculty, and relevant professional bodies to ensure that they are both legitimate and practically actionable. They should be accompanied by ongoing communication, regular revision cycles that track the rapidly evolving technological landscape, and an orientation that positions responsible AI use as a graduate competency to be developed rather than a boundary to be policed (Bacalso et al., 2026; Chan, 2023; Holmes et al., 2022; Tlili et al., 2024; UNESCO, 2023). Luckin, Holmes, Griffiths, and Forcier (2016) provides an influential model for such intelligence-augmentation governance frameworks, arguing that the goal of educational AI must be to enhance rather than replace human intelligence, a principle directly applicable to policy design for Generative AI in academic contexts.

7 Limitations and Future Research

This study has several limitations that qualify the scope of its conclusions and should be considered in the design of subsequent research.

The sample was drawn from institutions within a single national context. While the disciplinary diversity and the convergence of quantitative and qualitative findings across participant groups strengthen internal validity, the generalisability of the findings to higher education systems with different regulatory environments, academic cultural norms, or levels of digital infrastructure should be treated with caution. The cross-group universality documented in Section 6.5 increases the plausibility of transferability, but cross-national replication studies are necessary to test whether the Productivity–Cognitive Engagement Paradox and its structural characteristics are reproduced in contexts with different institutional AI governance landscapes.

The reliance on self-reported survey data introduces the possibility of social desirability bias. Students' responses on Interaction Transparency and Cognitive Engagement in particular may reflect aspirational rather than accurate self-assessments, given the social and institutional pressures associated with AI use disclosure. Future studies should complement survey instruments with objective behavioural data (such as learning management system logs, AI interaction histories collected with informed consent, or digital trace data) to provide direct behavioural validation of the self-reported patterns documented here.

The cross-sectional design precludes causal inference. The absence of significant regression relationships between individual-level predictors and Cognitive Engagement in the full sample is consistent with the structural paradox interpretation, but does not rule out causal relationships that operate at a different level of analysis (such as between-institution variation in policy and culture) or effects that would only emerge longitudinally. Longitudinal research designs tracking the same students across multiple semesters, measuring both dependency and engagement at repeated intervals, are necessary to determine whether the paradox intensifies, stabilises, or diminishes as students gain experience with Generative AI tools, and whether institutional policy changes produce discernible effects on engagement over time.

The qualitative corpus of 40 interview segments, while sufficient to achieve thematic saturation within the major identified categories, represents a constrained range of

voices and institutional contexts. The phenomenological findings are therefore appropriately understood as illustrative of structural dynamics rather than comprehensive accounts of the full diversity of student and faculty experience. Larger-scale qualitative investigations, including longitudinal ethnographic studies and institutional case studies tracing policy development and assessment transformation over time, would substantially enrich the empirical foundation of this area.

The study was conducted within a specific and rapidly evolving technological moment. The Generative AI tools available in 2025–2026 and the specific educational dynamics they produce may differ meaningfully from conditions arising as these technologies continue to develop. Findings should therefore be interpreted as characterising a particular phase of AI integration in higher education rather than as predictions holding as the technological landscape shifts.

The mediation analysis in this study employed the Sobel test to assess indirect effects. While appropriate here given the non-significant path coefficients, modern methodological practice favours bootstrapped confidence intervals (e.g., via Hayes' PROCESS macro) for their superior performance under non-normality and their direct estimation of effect-size intervals (Hayes, 2018). The Sobel approach is conservative and, in the present case, leads to the same conclusion; nonetheless, future replication work should adopt bootstrapped mediation to maximise comparability with contemporary standards.

Finally, the formal psychometric properties of the instruments, particularly construct-level internal consistency, require independent replication and confirmatory factor analysis validation before these scales are deployed as diagnostic tools in other research or institutional assessment contexts. The composite scores reported in this study should be treated as exploratory indices rather than fully validated psychometric measures. Construct validity and reliability evidence from larger, multi-site samples would substantially strengthen the foundation for the measurement approach proposed here, and this is an important priority for future instrument development work in the rapidly evolving field of AI integration in higher education.

8 Conclusion

This study provides convergent empirical evidence for the Productivity–Cognitive Engagement Paradox across contemporary higher education. Generative AI adoption has produced near-universal academic output gains while simultaneously displacing the cognitive engagement, critical thinking, and independent skill development that constitute genuine higher-order learning. The paradox is structurally robust: it is unmoderated by discipline, academic achievement, or usage intensity, and no individual-level characteristic predicts who is disengaged because the condition pervades the population rather than concentrating within identifiable subgroups.

The shadow usage phenomenon — systematic non-disclosure of AI involvement — reflects rational institutional responsiveness to policy ambiguity rather than integrity deficit. This reframing is consequential: it redirects appropriate intervention from punitive enforcement toward governance reform. On the faculty side, Institutional AI

Policy Clarity is the sole significant predictor of authentic task adoption, establishing that assessment transformation is enabled by institutional scaffolding rather than individual faculty seniority, initiative, or discipline.

The institutional policy vacuum, documented through both student shadow usage patterns and faculty trust ratings, operates not merely as an absence of guidance but as an active structural condition that amplifies the very behaviours it ostensibly seeks to govern. This study's contribution extends beyond documenting a problem; it provides the theoretical and empirical foundations for a coherent institutional response. The tripartite framework proposed here, integrating student AI literacy development, faculty-led assessment reform, and institution-level policy governance, offers a pathway from the current condition of productive but cognitively shallow education toward a future in which the genuine power of Generative AI is harnessed in service of, rather than at the expense of, deep learning and intellectual development.

Higher education institutions face a narrow and closing window in which to lead rather than merely react. The assessment transformation documented here is already underway at the faculty level; what is not yet in place is the broader pedagogical transformation, institutional coordination, and policy infrastructure necessary to govern it wisely. The empirical findings of this study constitute a call to action for institutional leaders, policymakers, curriculum designers, and educational researchers: an invitation to treat the integration of Generative AI into higher education not as an administrative compliance challenge but as one of the most consequential pedagogical questions of the contemporary period.

Declarations

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

Conflict of interest: The authors declare no conflicts of interest.

Ethics approval and consent to participate: Ethical approval was obtained from the relevant Institutional Review Board prior to data collection. All procedures were conducted in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments. Informed consent was obtained from all individual participants included in the study. All participant data are fully anonymised and stored in compliance with applicable data protection legislation.

Consent for publication: Not applicable. No identifiable participant information is included in this manuscript.

Data availability: The anonymised datasets analysed during the current study are available from the corresponding author upon reasonable request, subject to applicable data protection constraints.

Code availability: All Python analysis scripts used in this study are available from the corresponding author upon reasonable request.

Author contributions: [Withheld for double-blind peer review.]

References

Adiguzel, T., Kaya, M.H., Cansu, F.K. (2023). Revolutionizing education with AI:

Exploring the transformative potential of ChatGPT with special reference to Türkiye. *Contemporary Educational Technology*, *15*(3), ep429, https://doi.org/10.30935/cedtech/13152

Alhur, A.A., Khlaif, Z.N., Hamamra, B., Hussein, E. (2025). Paradox of AI in higher education: Qualitative inquiry into AI dependency among educators in Palestine. *JMIR Medical Education*, *11*, e74947, https://doi.org/10.2196/74947

Bacalso, F.D., Villamor, F.E., Paclipan, C.P., Orquia, C.J.B., Albiso, A.T., Himang, C.M. (2026). Institutional approaches to generative AI management in higher education: A systematic review. *Frontiers in Education*, *11*, , https://doi.org/10.3389/feduc.2026.1814426

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (ChatGPT): Understanding the potential benefits and challenges of ChatGPT in education. *SSRN Electronic Journal*, , https://doi.org/10.2139/ssrn.4337484

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bender, E.M., Gebru, T., McMillan-Major, A., Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 acm conference on fairness, accountability, and transparency* (pp. 610–623). New York: Association for Computing Machinery.

Biggs, J.B. (2003). *Teaching for quality learning at university* (2nd ed.). Buckingham: Open University Press.

Bittle, K., & El-Gayar, O.F. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, *16*(4), 296, https://doi.org/10.3390/info16040296

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101, https://doi.org/10.1191/1478088706qp063oa

Bretag, T. (2013). Challenges in addressing plagiarism in education. *PLOS Medicine*, *10*(12), e1001574, https://doi.org/10.1371/journal.pmed.1001574

Chan, C.K.Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, *20*, 38, https://doi.org/10.1186/s41239-023-00408-3

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology*, *32*(3), 444–452, https://doi.org/10.1007/s10956-023-10039-y

Cotton, D.R.E., Cotton, P.A., Shipway, J.R. (2023). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, *61*(2), 228–239, https://doi.org/10.1080/14703297.2023.2190148

Creswell, J.W., & Plano Clark, V.L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*, 22, https://doi.org/10.1186/s41239-023-00392-8

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y.-S., Gasevic, D., Chen, G. (2024). Can large language models provide feedback to students? a case study on ChatGPT. *IEEE Transactions on Learning Technologies*, *17*, 1850–1864, https://doi.org/10.1109/TLT.2024.3410427

Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319–340, https://doi.org/10.2307/249008

Deci, E.L., & Ryan, R.M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Springer.

Deci, E.L., & Ryan, R.M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268, https://doi.org/10.1207/S15327965PLI1104_01

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

Facione, P.A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction.* (California Academic Press, Millbrae, CA)

Farrokhnia, M., Banihashem, S.K., Noroozi, O., Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*, *61*(3), 460–474, https://doi.org/10.1080/14703297.2023.2195846

Firat, M. (2023). How ChatGPT can transform autodidactic experiences and open education. *Theory and Practice in Language Studies*, *13*(3), 718–722, https://doi.org/10.17507/tpls.1303.25

Fredricks, J.A., Blumenfeld, P.C., Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109, https://doi.org/10.3102/00346543074001059

Grassini, S. (2023). Shaping the future of education: Exploring the potential and consequences of AI and ChatGPT in educational settings. *Education Sciences*, *13*(7), 692, https://doi.org/10.3390/educsci13070692

Harris, C.R., Millman, K.J., van der Walt, S.J., et al. (2020). Array programming with NumPy. *Nature*, *585*, 357–362, https://doi.org/10.1038/s41586-020-2649-2

Hayes, A.F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York: Guilford Press.

Hmelo-Silver, C.E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review*, *16*(3), 235–266, https://doi.org/10.1023/B:EDPR.0000034022.16470.f3

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S.B., ... Koedinger, K.R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, *32*(2), 504–526, https://doi.org/10.1007/s40593-021-00239-1

Hon, K.K.L. (2026). Generative AI in higher education: A systematic review of its effects on learning outcomes and academic performance. *Journal of Educational Technology Systems*, , https://doi.org/10.1177/00472395251400089

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

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

Lim, W.M., Gunasekara, A., Pallant, J.L., Pallant, J.I., Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? a paradoxical perspective from management educators. *International Journal of Management Education*, *21*(2), 100790, https://doi.org/10.1016/j.ijme.2023.100790

Liu, Q., Hu, A., Gladman, T., et al. (2025). Eight months into reality: A scoping review of the application of ChatGPT in higher education teaching and learning. *Innovative Higher Education*, *50*, 1677–1700, https://doi.org/10.1007/s10755-025-09790-4

Lo, C.K. (2023). What is the impact of ChatGPT on education? a rapid review of the literature. *Education Sciences*, *13*(4), 410, https://doi.org/10.3390/educsci13040410

Luckin, R., Holmes, W., Griffiths, M., Forcier, L.B. (2016). *Intelligence unleashed: An argument for AI in education.* London: Pearson Education.

Mayer, R.C., Davis, J.H., Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734, https://doi.org/10.2307/258792

McCabe, D.L., Butterfield, K.D., Treviño, L.K. (2012). *Cheating in college: Why students do it and what educators can do about it.* Baltimore, MD: Johns Hopkins University Press.

McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th python in science conference* (pp. 56–61).

Mollick, E.R., & Mollick, L. (2023). *Assigning AI: Seven approaches for students, with prompts.* (SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4475995)

Ng, D.T.K., Leung, J.K.L., Chu, S.K.W., Qiao, M.S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041, https://doi.org/10.1016/j.caeai.2021.100041

OpenAI (2023). *GPT-4 technical report.* (arXiv preprint arXiv:2303.08774. https://arxiv.org/abs/2303.08774)

Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. *Computers and Education: Artificial Intelligence*, *2*, 100020, https://doi.org/10.1016/j.caeai.2021.100020

Perkins, M. (2023). Academic integrity considerations of AI large language models in education: A strategic approach. *Journal of University Teaching and Learning Practice*, *20*(7), 7, https://doi.org/10.53761/1.20.07.07

Python Software Foundation (2023). *Python language reference, version 3.12.* Retrieved from https://www.python.org

Qadir, J. (2023). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. *Proceedings of the 2023 ieee global engineering education conference (educon)* (pp. 1–9). IEEE.

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W.J., ... Marini, S. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, *6*(1), 41–56, https://doi.org/10.37074/jalt.2023.6.1.29

Revesai, Z. (2025). Generative AI dependency: The emerging academic crisis and its impact on student performance—a case study of a university in Zimbabwe. *Cogent Education*, *12*(1), 2549787, https://doi.org/10.1080/2331186X.2025.2549787

Ryan, R.M., & Deci, E.L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness.* New York: Guilford Press.

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare*, *11*(6), 887, https://doi.org/10.3390/healthcare11060887

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285, https://doi.org/10.1207/s15516709cog1202_4

Tlili, A., Abed, M., Bozkurt, A., et al. (2024). Navigating the ethical terrain of generative AI in education: A systematic review. *British Journal of Educational*

Technology, *55*(5), 1996–2017, https://doi.org/10.1111/bjet.13492

Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R., Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*, 15, https://doi.org/10.1186/s40561-023-00237-x

UNESCO (2023). *Guidance for generative AI in education and research.* (UNESCO, Paris. https://doi.org/10.54675/PCSP7720)

Virtanen, P., Gommers, R., Oliphant, T.E., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272, https://doi.org/10.1038/s41592-019-0686-2

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112, https://doi.org/10.1111/bjet.13370

Zawacki-Richter, O., Marín, V.I., Bond, M., Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, *16*, 39, https://doi.org/10.1186/s41239-019-0171-0

Appendix A Survey Constructs and Representative Items

Table A1: Student survey: selected representative items by construct

Construct	Representative Item
Perceived Usefulness	Using Generative AI tools improves the quality of my academic work.
Perceived Ease of Use	I find it straightforward to use Generative AI tools for academic purposes.
AI Literacy	I can evaluate the accuracy and limitations of AI-generated content.
Info Quality External (IQE)	AI-generated content is at least as reliable as information from conventional academic sources.
AI Dependency	I find it difficult to initiate academic tasks without first consulting an AI tool.
Cognitive Engagement	When completing assignments, I engage deeply with the underlying subject matter.
Critical Thinking	I independently evaluate the arguments and evidence relevant to my assignments.
Academic Productivity	Generative AI tools have significantly increased the volume of work I produce.
Interaction Transparency	I disclose my use of AI tools to my instructors when this is relevant to my work.

Table A2: Faculty survey: selected representative items by construct

Construct	Representative Item
Faculty Trust in Submissions	I am confident that standard student submissions reflect their own intellectual work.
Assessment Redesign	I have restructured my assessments specifically to address the challenges posed by AI.
Institutional AI Policy	My institution has provided clear and actionable guidance on acceptable AI use.
Pedagogical Transformation	My overall approach to teaching has changed significantly in response to AI adoption.