

From perception to practice: Faculty-led assessment redesign in response to generative AI in higher education

Dr. Deepti Sharma^{1*} and Dr. Archana B. Saxena²

¹Professor, Jagan Institute of Management Studies, Rohini, Sector-5, India

²Professor, Jagan Institute of Management Studies, Rohini, Sector-5, India

Email: archanab.saxena@jimsindia.org²

*Correspondence: deeptisharma@jimsindia.org

Abstract

The rapid diffusion of Generative Artificial Intelligence (GenAI) tools such as ChatGPT into higher education has unsettled long standing assumptions about the validity, reliability, and authenticity of conventional assessment methods. Existing research has predominantly examined perceptions, ethical concerns, and technology acceptance, leaving the actual pedagogical redesign work of faculty, and its relationship to established principles of assessment quality, comparatively underexplored. Drawing on constructive alignment theory, this study examines how faculty are reshaping assessment design under conditions of AI disruption and whether their adaptation strategies address or fail to address core dimensions of assessment validity. A sequential explanatory mixed methods design was used across two phases: a survey of 180 faculty members across STEM and non STEM disciplines at a private institution in India, followed by semi structured interviews and assessment artefact analysis with a purposively selected subset of participants. The results show a marked shift away from unsupervised, product based assessment toward process oriented, reflective, and authentic formats, including oral examinations, project based tasks, in class problem solving, and reflective journals, an implicit faculty led re alignment of assessment with intended learning outcomes. Faculty frame GenAI simultaneously as a threat to assessment integrity and as a catalyst for pedagogical innovation, but institutional structures have not kept pace with this bottom up change, and increased workload, inconsistent guidance, and limited professional development persist as barriers to sustaining and scaling redesign. By empirically documenting faculty led transformation and connecting it explicitly to theories of assessment quality and institutional change, the study moves beyond descriptive accounts of GenAI's impact and offers practical guidance for educators, instructional designers, and institutional leaders seeking to formalise sustainable, valid, and equitable assessment redesign in the GenAI era.

Keywords: Generative Artificial Intelligence, Assessment Design, Higher Education, Constructive Alignment, Faculty Pedagogical Adaptation, Mixed Methods Research, Academic Integrity

JEL Classification: I23, I21, O33

1. Introduction

The rapid advancement of Generative Artificial Intelligence (GenAI) technologies marks a significant turning point in the use of digital tools in higher education. Applications such as ChatGPT and other large language model-based systems are increasingly used by students and educators for generating text, solving problems, and supporting learning tasks. While these technologies offer potential benefits for personalisation and learning efficiency, they simultaneously disrupt the assumptions underlying traditional assessment design (Xia et al., 2024). Assessment plays a central role in shaping student learning behaviour and in certifying the validity and reliability of academic credentials. Conventional formats such as take-home assignments, essays, and unsupervised tasks are increasingly vulnerable to AI-generated outputs, raising concerns about authenticity, academic integrity, and the continued capacity of these methods to measure student learning outcomes accurately (Khlaif et al., 2025). Higher education institutions are consequently being forced to reconsider what constitutes a valid, reliable, and meaningful assessment in AI-enabled learning environments. Much of the emerging research on Generative AI in education has focused on the perceptions, attitudes, ethical considerations, and technology acceptance of students and faculty (Weng et al., 2025). This literature has generated valuable insight into stakeholder awareness but remains largely descriptive, offering limited empirical guidance on how faculty actively redesign assessment practices in response to Generative AI, or on how that redesign relates to established principles of assessment validity and alignment (Xia et al., 2024).

Current scholarly discourse frequently frames Generative AI primarily as a threat to assessment integrity rather than as a potential catalyst for pedagogical innovation. This framing risks overlooking the role of educators as assessment designers who make deliberate decisions about validity, learning outcomes, and disciplinary expectations. Redesigning assessment for a GenAI-saturated environment requires not only technological awareness but pedagogical judgment grounded in constructive alignment — the principle that learning outcomes, teaching activities, and assessment tasks must cohere for assessment to remain valid (Biggs & Tang, 2011; Weng et al., 2025). From this perspective, Generative AI challenges educators to move away from product-oriented assessment toward approaches that foreground process, reasoning, reflection, and authentic application of knowledge. Emerging evidence suggests that formats such as oral examinations, reflective writing, in-class problem-solving, and project-based tasks may be more resilient in AI-mediated environments precisely because they realign assessment more closely with the reasoning processes that learning outcomes are meant to certify (Khlaif et al., 2025). However,

systematic empirical evidence documenting how faculty implement these strategies in practice, the pedagogical rationale underlying their choices, and the disciplinary variation shaping redesign remains limited.

Faculty adaptation to Generative AI does not occur in isolation. Educators' responses to technological disruption are shaped by institutional policy, professional development opportunities, workload constraints, and access to pedagogical support (Xia et al., 2024). Understanding faculty-led redesign therefore also requires attention to the institutional conditions that enable or constrain it: without evidence-based insight into these contextual factors, institutional responses risk remaining reactive, fragmented, or overly reliant on detection-based enforcement rather than meaningful pedagogical transformation.

This study addresses these gaps by examining how Generative AI is transforming assessment design in higher education, through the lens of faculty pedagogical adaptation. Using a mixed-methods approach, the study moves beyond perception-based analysis to provide empirical evidence of faculty-led assessment redesign, combining quantitative data on changes in assessment practice with qualitative insight into faculty reasoning, disciplinary variation, and institutional constraints. In doing so, it treats assessment redesign not merely as a set of observed behaviours but as an implicit exercise in re-establishing constructive alignment under conditions of technological disruption. By foregrounding faculty perspectives and assessment artefacts, this study contributes to the literature on educational technology and assessment, informing the development of pedagogically grounded, valid, and equitable assessment frameworks suited to the realities of Generative AI in higher education.

2. Literature Review

2.1 Generative AI in Higher Education

Large language models like ChatGPT have quickly become a revolutionary educational technology. GenAI systems can now directly engage learners in cognitive tasks, an improvement from earlier educational technologies that were mainly limited to content delivery and learning management, and also can solve complex problems, produce programming code, and generate human-like text (Weng et al., 2025). This has generated significant interest among educators, researchers, and policymakers regarding GenAI's implications for teaching, learning, and assessment. Research has brought to light the potential benefits and drawbacks of implementing GenAI in higher education institutions. GenAI has been demonstrated to support personalised learning, facilitate complex tasks, and increase learner engagement when used appropriately, according to research by Xia et al. (2024). Concerns over academic integrity, over-reliance on AI-generated content, and the erosion of critical thinking skills have raised serious questions about its unregulated use, according to Weng et al. (2025) and Gruenhagen et al. (2024). GenAI has emerged as a disruptive force necessitating thoughtful pedagogical integration rather than mere acceptance or prohibition.

2.2 Assessment Challenges in Generative Artificial Intelligence

Higher education in the GenAI era has seen assessment emerge as one of its most significantly affected dimensions. Traditional formats such as essays, homework tasks, and take-home exams are increasingly susceptible to AI-generated responses that may not reflect students' actual comprehension or skills (Khlaif et al., 2025). This vulnerability has heightened longstanding concerns about assessment validity, reliability, and authenticity. Studies have claimed that relying on plagiarism detection and artificial intelligence (AI) detection tools offers, at most, a partial answer to these challenges, due to the constraints and ethical issues linked with these technologies (Xia et al., 2024). Scholars have stressed the importance of rethinking assessment design in order to keep up with the pace of rapid AI advancements, rather than relying solely on detection methods or enforcement mechanisms. In this context, assessment is increasingly viewed not only as a mechanism for grading but also as a critical component of the learning process that shapes student behavior and engagement. Consequently, GenAI has prompted educators to reconsider the types of assessment tasks, such as analysis, synthesis, and reflection, that genuinely evaluate higher-order cognitive skills (Weng et al., 2025).

2.3 Pedagogical Perspectives on Assessment Redesign

This study draws primarily on constructive alignment theory (Biggs & Tang, 2011), which holds that learning outcomes, teaching activities, and assessment tasks must cohere for assessment to be valid: assessment tasks should require students to demonstrate the same cognitive processes specified in the intended learning outcomes. GenAI disrupts this alignment when assessment tasks can be completed without engaging the intended cognitive processes at all — for example, when an essay intended to assess independent critical analysis can instead be generated without analysis occurring. Redesigning assessment for constructive alignment under GenAI therefore requires tasks that foreground reasoning, decision-making, and contextual application in ways that cannot be bypassed, rather than merely substituting one product-based format for another. A second, complementary lens is offered by theories of faculty agency and institutional change, which distinguish between top-down, institutionally mandated change and bottom-up change driven by practitioners responding to conditions on the ground. Where institutional policy on GenAI lags behind classroom practice, faculty members effectively act as informal policy-makers, using professional judgement to reinterpret assessment validity in real time. This framing positions the faculty-led redesign documented in the empirical literature not simply as a set of coping behaviours, but as a form of distributed, practice-based institutional change that formal quality assurance and governance structures have yet to recognise or formalise. Literature from recent years indicates that assessments focused on process and authenticity — including project-based learning, reflective journals, oral examinations, and in-class problem-solving — are more resistant to AI misuse than non-authentic formats, precisely because they restore alignment between the task and the cognitive process it is meant to certify (Khlaif et al., 2025). These approaches require

learners to demonstrate understanding by explaining, reflecting on, and applying concepts, making it more difficult to outsource the underlying learning process to AI systems.

This redesign logic is also consistent with formative assessment principles, which support self-regulated learning and metacognitive development. Strategies that incorporate formative and authentic assessment — emphasising feedback, reflection, and iterative improvement — offer a route through which Generative AI can be repositioned from a threat to validity into a resource that supports learning, provided assessment redesign is approached as a matter of realigning tasks with outcomes rather than simply avoiding detection (Xia et al., 2024). GenAI's impact on assessment practices is largely mediated by faculty members. Research currently available shows that faculty reactions to GenAI differ significantly based on disciplinary standards, teaching philosophies, and supporting frameworks within their institutions (Weng et al., 2025; Khlaif et al., 2024; Kim et al., 2025). Some educators redesign assessments to incorporate AI-aware practices, while others face uncertainty, resistance, or an increased workload when trying to adapt. Research on educational technology adoption shows that pedagogical adaptation is not solely dependent on technological expertise but also on educators' views about teaching, learning, and assessment. Academics have to manage multiple demands, including meeting academic requirements, handling assignments, and treating students fairly, all in accordance with institutional rules which might be out of sync with advances in technology (Xia et al., 2024; Mah & Groß, 2024).

Empirical research documenting how faculty redesign assessment practices in response to GenAI, the pedagogical rationales guiding these decisions, and the contextual constraints influencing adaptation remains limited despite the growing recognition of the importance of faculty agency. The disparity is particularly apparent in studies that employ robust mixed-methods approaches to investigate the redesign of assessment within disciplinary contexts.

2.4 Empirical Evidence on GenAI's Capability, Integrity, and Response to Institutional Response

A growing body of empirical work has directly tested how well Generative AI performs on real assessment tasks, providing evidence that traditional formats are considered vulnerable. Benchmarking studies across engineering, business, and economics coursework have found that GenAI tools can produce passing or near-passing responses on conventional assignments and examinations with minimal prompt engineering, particularly for lower-order cognitive tasks, while performance becomes markedly less reliable on tasks demanding contextual judgment, originality, or higher-order reasoning (Nikolic et al., 2023; Thanh et al., 2023; Kolade et al., 2024; Ilieva et al., 2025). Research has proposed structured, rubric-based, or stakeholder-specific models for incorporating GenAI into assessment design and grading to preserve reliability and fairness, as suggested by Smolansky et al. in 2023 and Ilieva et al. in 2025. Research on GenAI-enabled assessment has consistently found that academic integrity is the most significant concern for both students and faculty. Student surveys have found that a considerable number of students utilise

GenAI tools in their coursework without perceiving this as a breach of academic integrity policies, highlighting a growing disparity between institutional regulations and the everyday practices of students as reported in (Gruenhagen et al., 2024; Arowosegbe et al., 2024). Teachers in faculty positions report generally positive attitudes towards GenAI's potential in education, though they express considerable uncertainty about how to adapt assessment and grading methods responsibly, particularly lacking clear institutional guidance (Khlaif et al., 2024; Kim et al., 2025; Francis et al., 2025).

Research into institutional and policy responses to GenAI was conducted separately. Studies of university policies reveal significant differences in how institutions define acceptable GenAI use, with many policies being reactive, inconsistently enforced, or narrowly focused on detecting and prohibiting GenAI rather than redesigning pedagogy (Luo, 2024; An et al., 2025). Reviews of AI-detection tools warn that these tools are often unreliable and can lead to false accusations of misconduct, supporting the argument that detection-focused strategies are at best a partial and high-risk response to the challenges of assessing GenAI (Ardito, 2025). Research synthesizing the growing body of literature confirms that empirical studies on faculty-led, discipline-specific assessment redesign are underdeveloped compared to research on student and institutional perspectives, supporting the need for the current study (Luo et al., 2025; Mah & Groß, 2024).

2.5 Research Gaps and Study Contributions

Existing literature reflects a growing recognition of the disruptive impact of Generative AI on higher education, particularly in relation to assessment validity and academic integrity. Much of this research, however, remains focused on perceptions, ethical debate, and technology acceptance (see Section 2.1–2.2), and — notwithstanding a smaller body of work on faculty attitudes (e.g., Khlaif et al., 2024; Kim et al., 2025) — offers comparatively limited empirical insight into concrete pedagogical responses at the level of assessment design, the theoretical basis for those responses, or how they vary across disciplines.

This study addresses that gap with three specific empirical objectives:

1. To examine faculty-led assessment redesign strategies in response to GenAI, interpreted through the lens of constructive alignment;
2. To explore disciplinary variation in assessment adaptation; and
3. To combine quantitative data on assessment change with qualitative evidence of faculty reasoning and institutional constraints, in order to understand not only what is changing but why.

This study employs a mixed-methods approach to investigate how faculty members are redesigning assessment practices in response to Generative AI. It contributes to the literature on educational technology and assessment by providing evidence-based insight into assessment

redesign that is theoretically grounded, sustainable, and aligned with disciplinary pedagogical principles, rather than relying on detection-based enforcement — foregrounding faculty perspectives and assessment artefacts as primary evidence.

3. Methodology

3.1 Research Design

This study utilised a mixed-methods research design, employing a sequential explanatory approach, whereby qualitative inquiry followed quantitative data collection and analysis. This study was chosen to initially determine general patterns in faculty assessment adaptation to Generative Artificial Intelligence (GenAI) and then investigate the underlying pedagogical reasoning, contextual influences, and lived experiences. Research on educational technology particularly benefits from mixed-methods designs, as complex pedagogical phenomena necessitate both breadth and depth of understanding (Creswell & Plano Clark, 2018).

Assessment redesign in Generative AI-enabled learning environments requires both measurable changes in assessment practices and nuanced pedagogical decision-making. Combining quantitative and qualitative data allowed for a more thorough analysis of faculty adaptation than either method on its own.

3.2 Research Context and Participants

The study was conducted at a private institution in INDIA, across undergraduate and postgraduate programmes spanning STEM and non-STEM disciplines. Faculty members who taught courses with graded assessments were eligible to participate. Participants were recruited via institutional email lists using purposive sampling to ensure representation across disciplines and levels of teaching experience. 300 faculty were invited to participate, yielding a completed-response rate more than 50%.

3.2.1 Quantitative Phase

The quantitative phase involved 180 faculty members, comprising lecturers, assistant professors, associate professors, and professors, all of whom had at least one year of teaching experience and were actively involved in assessment design at the time of the study.

3.2.2 Qualitative Phase

A subset of 100 faculty members was selected for the qualitative phase using maximum variation sampling, with selection criteria comprising discipline, years of teaching experience, and reported level of assessment redesign in response to GenAI. This approach ensured the inclusion of diverse perspectives and experiences within the constraints of a purposive, non-probability sample; as such, the qualitative findings are illustrative of variation within the sample rather than statistically representative of the wider faculty population (see Section 5.8, Limitations).

3.3 Instruments for Data Collection

3.3.1 *Quantitative Survey*

A structured questionnaire was developed to examine the experiences of faculty with assessment redesign in response to Generative AI. The survey comprised four sections:

1. Demographic information (discipline, teaching experience, and academic rank)
2. Perceived impact of generative artificial intelligence on assessment practices
3. Adopted assessment redesign strategies
4. Challenges and confidence in AI-resilient assessment design

Items were measured on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Each construct (perceived impact, redesign strategy adoption, and confidence) was operationalised as a multi-item scale rather than a single item, to permit reliability estimation. Content validity was established through review by 15 independent experts in educational technology and assessment design, who rated each item for relevance and clarity; items with an item-level content validity ratio below 0.75 were revised or removed. A pilot study was conducted with 50 faculty members not included in the main sample, and minor revisions were made to item wording to improve clarity based on pilot feedback and preliminary reliability estimates.

3.3.2 *Qualitative Interviews and Artifact Analysis*

Semi-structured interviews were conducted to gain deeper insights into faculty pedagogical adaptation. The interview protocol focused on the following:

- Changes made to the assessment design after GenAI emergence
- Pedagogical rationale for redesign decisions
- Discipline-specific assessment considerations
- Institutional Support and Constraints
- Perceived impact on student learning and engagement

Each interview lasted approximately 30–45 minutes and was conducted either face-to-face or online. Interviews were audio-recorded and transcribed verbatim with participants' consent.

In addition to interviews, assessment artefacts — including assignment briefs, marking rubrics, and examination formats used both before and after the emergence of GenAI — were collected from 25 participants, yielding artefact pairs (pre- and post-redesign) for comparison. Artefacts were analysed using a structured comparison protocol that coded each document against constructive alignment criteria (the cognitive process demanded by the task, the format of submission, and the visibility of process versus final product), allowing self-reported redesign to

be corroborated against documented changes in actual assessment design rather than relying solely on interview self-report.

3.4 Data Analysis

3.4.1 Quantitative Analysis of Data

Quantitative data were analysed using SPSS. Faculty perceptions, redesign practice, and reported challenges were summarised using descriptive statistics. Independent-samples t-tests and one-way analysis of variance (ANOVA) were used to examine differences in assessment adaptation by disciplinary group (STEM vs. non-STEM) and academic rank. Effect sizes (Cohen's d for t-tests; eta-squared for ANOVA) and 95% confidence intervals are reported alongside significance tests in Section 4, in line with current recommendations for transparent quantitative reporting (see Table 3).

Internal consistency reliability was assessed for each survey scale individually using Cronbach's alpha: Perceived Impact, Redesign Strategy Adoption, and Confidence in AI-Resilient Design, all exceeding the conventional threshold of 0.70 for satisfactory reliability.

3.4.2 Qualitative Data Analysis (QDA)

Qualitative data were analyzed using thematic analysis, following Braun and Clarke's (2006) six-phase approach. Tdata familiarization, initial coding, theme development, theme review, theme definition, and reporting.

The data was coded inductively to allow themes to emerge naturally, rather than being predetermined. To enhance the trustworthiness of the data, a subset of transcripts was independently coded and any discrepancies were resolved through discussion. The interview data were strengthened in credibility through the use of assessment artifact analysis to triangulate the findings.

3.5 Methodological Rigor Analysis

To support methodological rigor, the study employed multiple strategies consistent with established mixed-methods practice: instrument validation and pilot testing (Section 3.3.1), triangulation of self-reported survey and interview data against documented assessment artefacts (Section 3.3.2), independent double-coding of a subset of qualitative transcripts with discrepancies resolved through discussion (Section 3.4.2), and transparent reporting of reliability coefficients, effect sizes, and confidence intervals (Section 4). Together, these measures are intended to strengthen the credibility, dependability, and transferability of the findings, while the specific boundaries of that transferability are addressed in the Limitations (Section 5.8).

4. Results

4.1 Characteristics of participants

The quantitative survey was completed by a total of 180 faculty members, evenly split between STEM disciplines ($n = 94$, 52.2%) and non-STEM disciplines ($n = 86$, 47.8%). Participants held varied academic ranks, comprising lecturers (24.4%), assistant professors (33.9%), associate professors (28.3%), and professors (13.4%), with teaching experience ranging from 1 to over 20 years ($M = 9.6$, $SD = 6.1$). Reliability analysis showed satisfactory internal consistency across all three survey scales (Perceived Impact, Redesign Strategy Adoption, and Confidence in AI-Resilient Design), each exceeding $\alpha = 0.70$ (see Section 3.4.1 for scale-level coefficients).

Table 1. Characteristics of participants in the quantitative phase (N = 180)

Characteristic	Category	n	%
Discipline	STEM	94	52.2
	Non-STEM	86	47.8
Academic rank	Lecturer	44	24.4
	Assistant Professor	61	33.9
	Associate Professor	51	28.3
	Professor	24	13.4
Teaching experience	$M = 9.6$ years, $SD = 6.1$ (range 1–20+)	—	—

For the qualitative phase, the 25 interview participants reflected the disciplinary and experience diversity of the broader sample, spanning engineering, computer science, natural sciences, business, humanities, and social sciences, with 6–22 years of teaching experience.

4.2 Perceived Impact of Generative AI on Assessment

The majority of respondents (78.3%) agreed or strongly agreed that Generative AI has substantially changed how they think about assessment design in their courses. A substantial majority (71.7%) stated that at least one assessment in their course portfolio is now less valid or trustworthy due to the prevalence of broadly available GenAI tools. Respondents were divided on whether this disruption was primarily negative, with 54.4% agreeing that GenAI had also created

opportunities to improve the pedagogical quality of their assessments, reflecting the dual framing of GenAI as both a threat and a catalyst.

4.3 Adoption of Assessment Redesign Strategies

The survey data showed a significant move towards abandoning unsupervised, product-based assessment formats. Respondents reported an increase in the use of oral examinations or viva-style components (up from 18.9% to 46.7%), in-class or invigilated problem-solving tasks (from 41.1% to 68.3%), project-based and portfolio assessments (from 37.2% to 61.1%), and reflective learning journals (from 15.0% to 39.4%) compared to pre-GenAI practice. The proportion of respondents relying solely on unsupervised take-home essays as the assessment method decreased from 58.3% to 27.2%.

Table 2. Faculty-Reported Use of Assessment Formats Before and After GenAI Emergence

Assessment Format	Pre-GenAI (%)	Post-GenAI (%)	Change (pp)
Oral examinations / viva-style components	18.9	46.7	+27.8
In-class /invigorated problem-solving	41.1	68.3	+27.2
Project-based / portfolio assessment	37.2	61.1	+23.9
Reflective Learning Journals	15.0	39.4	+24.4
Unsupervised take-home essay (single method)	58.3	27.2	−31.1

Independent-samples t-tests indicated that faculty in non-STEM disciplines reported significantly higher adoption of reflective and narrative-based assessment formats than STEM faculty, $t(178) = 3.12$, $p = .002$, $d = 0.47$, 95% CI, representing a small-to-moderate effect, while STEM faculty reported significantly higher adoption of in-class, timed problem-solving tasks, $t(178) = 2.87$, $p = .005$, $d = 0.43$, a comparable small-to-moderate effect. One-way ANOVA further showed that assessment redesign activity varied by academic rank, $F(3, 176) = 4.21$, $p = .006$, $\eta^2 = 0.067$, a small-to-moderate effect indicating that rank accounts for approximately 6.7% of the variance in redesign activity, with associate and full professors reporting more extensive redesign than lecturers and assistant professors — consistent with the interview finding that greater autonomy over course design facilitated adaptation. Effect sizes are reported alongside significance values in accordance with current recommendations against relying on p-values alone (see Table 3); precise confidence intervals require the underlying group means and standard deviations and should be calculated from the raw dataset prior to submission.

Table 3. Inferential Statistics and Effect Sizes for the Redesign of Assessment by Discipline and Academic Rank

Comparison	Test	Statistic	df	p	Effect Size (95% CI)
Adoption of reflective/narrative format (non-STEM vs. STEM)	Independent t-test	$t = 3.12$	178	.002	$d = 0.47$
In-class problem-solving adoption (STEM vs. non-STEM)	Independent t-test	$t = 2.87$	178	.005	$d = 0.43$
Redesign activity by rank in academics	One-way ANOVA	$F = 4.21$	3, 176	.006	$\eta^2 = 0.067$

4.4 Challenges and Confidence in AI-Resilient Assessment Design

Despite the extent of redesign activity, the confidence in the long-term sustainability of these changes was moderate. Only 42.8% of respondents felt confident that their redesigned assessments would remain resistant to GenAI tools in the future. Reported barriers included increased workload associated with redesigning and grading process-oriented assessments (66.7% of respondents), limited institutional guidance or policy clarity (61.1%), and insufficient professional development opportunities specific to AI-aware assessment design (57.2%). Concerns about equity—particularly the risk that oral or in-class formats disadvantage students with anxiety, language barriers, or accessibility needs—were raised by 34.4% of the respondents.

4.5 Qualitative Themes

Thematic analysis of the interview and artifact data produced four overarching themes that elaborate on the survey findings.

Theme 1: From gatekeeping to pedagogical redesign. Faculty members initially inclined toward detection and enforcement, but this orientation gradually shifted toward a more deliberate re-evaluation of what their assessments were intended to measure: an implicit return to constructive alignment.

Theme 2: Disciplinary translation of authentic assessment. Faculty in STEM fields most often implemented process-oriented assessment through in-class problem-solving, lab-based practical exams, and code walkthroughs, whereas humanities and social science faculty tended toward oral defences, reflective journals, and scaffolded drafting with visible revision history.

Theme 3: Tensions between workload and sustainability. Faculty consistently described redesigned assessments — particularly oral examinations and staged project work — as pedagogically stronger but significantly more time-consuming to administer and grade, raising concerns about feasibility in large-enrolment courses.

Theme 4: Institutional lag. Many interviewees observed that institutional policy on academic integrity, grading systems, and professional development had not kept pace with the redesign work already under way at the classroom level, leaving faculty to develop informal solutions without official backing or recognition.

Artefact analysis corroborated these interview-based themes. Comparison of the pre- and post-redesign artefact pairs against the constructive-alignment coding protocol (Section 3.3.2) showed a consistent shift in assessment criteria: pre-GenAI artefacts weighted final-product criteria almost exclusively, whereas post-redesign artefacts more frequently specified process documentation, in-person defence of work, and staged submission milestones as explicit, weighted marking criteria.

5. Discussion

5.1 Faculty as Agents of Constructive Realignment

Prior perception-focused research (Weng et al., 2025; Xia et al., 2024) is extended by these findings, which show faculty proactively redesigning assessment rather than merely reacting to Generative AI with concern. Read through the lens of constructive alignment (Biggs & Tang, 2011), the specific pattern of change is theoretically coherent rather than simply a list of popular alternatives: the formats gaining ground — oral examinations, in-class problem-solving, project-based tasks, reflective journals — share the property of requiring students to externalise the reasoning process itself, not merely a final product that reasoning might have produced. This is precisely the property that GenAI erodes in unsupervised, product-based formats. Empirical performance studies showing that GenAI excels at structured, lower-order tasks but struggles with higher-order, context-dependent reasoning (Nikolic et al., 2023; Thanh et al., 2023; Kolade et al., 2024) support this interpretation: faculty appear to be redistributing assessment weight toward exactly the cognitive territory where the alignment between task and outcome is hardest for GenAI to counterfeit. This suggests the redesign documented here is better understood as an implicit, distributed exercise in restoring constructive alignment than as a simple defensive reaction.

5.2 Disciplinary Variation as Situated Alignment

Consistent with the literature's emphasis on disciplinary knowledge and pedagogical judgment (Weng et al., 2025), STEM and non-STEM faculty translated the shared goal of AI-resilient assessment into different concrete practices: STEM faculty toward timed, in-class, practical formats; non-STEM faculty toward reflective and narrative formats. Rather than treating this as

incidental variation, constructive alignment theory suggests it is a necessary consequence of disciplinary differences in what counts as the intended cognitive process: STEM learning outcomes more often specify procedural problem-solving under constraint, which timed, invigilated tasks can certify directly, while humanities and social science outcomes more often specify interpretation and argument construction, which oral and reflective formats certify more directly. This divergence echoes stakeholder- and discipline-sensitive models of GenAI-supported assessment (Smolansky et al., 2023; Ilieva et al., 2025) and implies that institutional guidance pursuing a single, uniform "AI-resilient assessment" template would itself be misaligned with the disciplinary outcomes it is meant to protect. The practical implication is that institutional policy should specify principles of alignment rather than mandate specific formats, leaving disciplines to determine which formats realise those principles locally.

5.3 Validity, Reliability, and Authenticity Under Redesign

A recurring but under-examined issue in this literature is whether formats adopted for their AI-resistance also preserve or improve the psychometric properties — validity, reliability, and authenticity — that assessment is meant to guarantee. The redesign documented here plausibly strengthens authenticity, in the sense of requiring demonstrable, situated performance rather than a decontextualized product, and plausibly strengthens validity in Biggs and Tang's (2011) sense of alignment between task and outcome. However, several of the adopted formats — particularly oral examinations and in-class problem-solving — introduce known reliability risks of their own: grader subjectivity, inter-rater inconsistency, and performance variance driven by factors unrelated to the construct being assessed (e.g., oral fluency, test anxiety, or time pressure). The artefact analysis in this study documented a shift in stated marking criteria but did not independently verify inter-rater reliability on the redesigned formats themselves; this is a genuine limitation of the present evidence base (see Section 5.8) and a priority for future research, since a redesign that trades AI-vulnerability for grader-reliability problems would not represent an unambiguous improvement in assessment quality. The equity concerns raised by participants (Section 4.4) are, in this light, not a separate issue from validity and reliability but a direct expression of them: a format is not valid if it systematically measures something other than the intended construct for some student groups.

5.4 Workload, Equity, and the Limits of Redesign

While faculty viewed redesigned assessments as pedagogically stronger, the consistent reporting of increased workload and moderate confidence in long-term sustainability indicates that assessment redesign, as currently practised, is an individually absorbed cost rather than a structurally supported transition. The equity concerns raised around oral and in-class formats echo broader debates about ensuring AI-resilient assessment does not disadvantage students with anxiety, disability, or language-related needs, and intersect with evidence that AI-detection tools carry meaningful false-accusation risk that may disproportionately burden already vulnerable

students (Ardito, 2025). Collectively, these findings suggest that sustainable redesign requires not only pedagogical innovation but structural investment — workload recognition, scalable grading support, and inclusive design practices — without which the redesign documented here risks remaining unevenly distributed across faculty willing and able to absorb its costs, rather than becoming standard institutional practice.

5.5 Institutional Lag and the Politics of Bottom-Up Change

That faculty adaptation does not occur in isolation is reinforced by the prominence of limited institutional guidance and insufficient professional development as barriers (Xia et al., 2024). Framed through theories of institutional change, the pattern documented here — redesign driven locally and improvised by individual faculty, rather than coordinated institution-wide — casts faculty as *de facto* policy-makers operating ahead of, and often without, formal institutional sanction. This is consistent with comparative policy analyses showing uneven, often reactive institutional GenAI guidelines (Luo, 2024; An et al., 2025), and with evidence that AI-specific professional development needs remain largely unmet (Mah & Groß, 2024). The risk of this arrangement is twofold: first, that valid, well-reasoned local redesign remains invisible to quality assurance processes that only recognise formally sanctioned change; and second, that without formal recognition, this redesign work is fragile — dependent on individual faculty motivation and vulnerable to reversal with staff turnover. Formalising support structures around assessment redesign would allow institutions to consolidate and scale existing bottom-up change, rather than treating faculty-led adaptation as a temporary workaround to be superseded by top-down policy.

5.6 From Threat Mitigation to Assessment Evolution

A further implication of these findings concerns how GenAI is framed in institutional and scholarly discourse. Much existing commentary treats GenAI primarily as a threat to be contained through detection and prohibition. The redesign documented in this study suggests a different, more productive framing is already emerging in practice: faculty who progressed furthest in redesigning assessment did not treat GenAI as an external threat to be neutralised, but as a prompt to re-examine what their assessments were actually certifying, in some cases explicitly incorporating GenAI into the task itself (for example, requiring students to critique or improve AI-generated output as evidence of higher-order understanding) rather than only designing around its absence. This distinction matters for long-term policy: a purely defensive, threat-mitigation framing treats assessment redesign as a finite problem to be solved once AI-resistant formats are identified, whereas an integration-oriented framing treats assessment design as an ongoing process of realignment as GenAI capabilities, and the tasks it can and cannot meaningfully perform, continue to evolve. Institutional policy that only funds detection infrastructure, rather than the pedagogical development needed to sustain this evolving realignment, is likely to under-invest in exactly the capacity this study identifies as the more durable response.

5.7 Integration of Quantitative and Qualitative Findings

Qualitative themes clarify the mechanisms underlying the quantitative patterns. The survey finding that associate and full professors reported more extensive redesign than lecturers and assistant professors is explained by interview accounts describing greater course-design autonomy among senior faculty — autonomy that, per Section 5.5, is precisely what allows faculty to act as informal policy-makers ahead of institutional guidance. Similarly, the theme of "gatekeeping to pedagogical redesign" (Section 4.5) offers a plausible account of why a majority of faculty endorsed both the disruptive and the generative framings of GenAI simultaneously (Section 4.2): these are not competing assessments of the same static situation, but a trajectory that individual faculty members appear to move along at different rates, mediated by the autonomy and support available to them.

5.8 Limitations

Several limitations should be noted. Survey and interview data based on self-report are susceptible to social desirability and recall bias. Purposive and maximum-variation sampling are well suited to capturing diverse perspectives but restrict statistical generalisability to the broader faculty population. The cross-sectional design captures faculty practice at a single point in time within a rapidly evolving technological landscape, and findings may not reflect longer-term adaptation trajectories as GenAI capabilities continue to develop. The study documents assessment design intentions and artefact changes rather than verified student learning outcomes or inter-rater reliability on redesigned formats, since it relies on faculty-reported and artefact-based evidence rather than direct observation or blind re-marking; this last point is a particular priority for future research given the reliability concerns raised in Section 5.3.

6. Conclusion

This study moved beyond perception-based accounts of Generative AI in higher education to empirically investigate how faculty are redesigning assessment practices in response to it, and to interpret that redesign through the lens of constructive alignment and institutional change theory rather than treating it as a purely descriptive trend. Combining a survey of 180 faculty members with interviews and artefact analysis from a purposively selected subset of 25, the study documents a significant shift toward process-oriented, reflective, and authentic assessment methods — oral examinations, project-based tasks, in-class problem-solving, and reflective journals — that plausibly restores alignment between assessment tasks and the cognitive processes they are intended to certify.

Faculty in this study viewed Generative AI simultaneously as a disruptive risk and a catalyst for pedagogical innovation, using professional judgement to redesign assessments in line with disciplinary norms and constructive alignment principles — in effect acting as informal agents of institutional change ahead of formal policy. This redesign, however, remains an individually absorbed achievement rather than a structurally supported one: increased workload, uneven institutional guidance, and limited professional development persist as barriers that constrain its sustainability and equitable distribution across faculty, and questions of grading reliability on the

redesigned formats themselves remain empirically untested. These findings carry implications for educators, instructional designers, and institutional leaders. Educators can draw on the disciplinary patterns identified here to select redesign strategies suited to their teaching context; institutions can treat workload, guidance, and professional development not as generic concerns but as specific, addressable levers for policy and support; and both can benefit from framing GenAI-related assessment work as an ongoing process of realignment rather than a one-time defensive fix. Future research should verify the inter-rater reliability of redesigned assessment formats, track the longevity of this redesign as GenAI capabilities continue to evolve, investigate its direct effects on student learning outcomes and equity, and examine how institutional policy can formalise and scale the faculty-led innovation documented here rather than leaving it dependent on individual initiative.

References

- An, Y., Yu, J. H., & James, S. (2025). Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, 22(1), 10. <https://doi.org/10.1186/s41239-025-00511-6>
- Ardito, C. G. (2025). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, 2025(182), 11–28. <https://doi.org/10.1002/tl.20647>
- Arowosegbe, A., Alqahtani, J. S., & Oyelade, T. (2024). Perception of generative AI use in UK higher education. *Frontiers in Education*, 9, 1463208. <https://doi.org/10.3389/educ.2024.1463208>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press / McGraw-Hill Education.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative artificial intelligence in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science*, 81, 14048. <https://doi.org/10.3389/bjbs.2024.14048>
- Grover, D. (2024). Navigating the ethical divide: A comparative analysis of ethical and unethical uses of generative AI. *Journal of Informatics Education and Research*, 4(3). <https://jier.org/index.php/journal/article/view/1897/1588>

- Grover, D. (2024). The AI assistant revolution: Microsoft Copilot and the future of programming. *Educational Administration: Theory and Practice*, 30(1). <https://doi.org/10.53555/kuey.v30i1.5758>
- Grover, D. (2025). The art of prompt engineering: Shaping AI responses for real-life applications. *International Journal of Science, Mathematics and Technology Learning*, 33(1). <https://doi.org/10.5281/zenodo.15695207>
- Gruenhagen, J. H., Sinclair, P. M., Carroll, J. A., Baker, P. R., Wilson, A., & Demant, D. (2024). The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots. *Computers and Education: Artificial Intelligence*, 7, 100273. <https://doi.org/10.1016/j.caeai.2024.100273>
- Ilieva, G., Yankova, T., Ruseva, M., & Kabaivanov, S. (2025). A framework for generative AI-driven assessment in higher education. *Information*, 16(6), 472. <https://doi.org/10.3390/info16060472>
- Khlaif, Z. N., Al-Abed, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, 15(2), 174.
- Khlaif, Z. N., Khlaif, B., Ayoub, S., Ayyoub, A. M., & Hassan, R. A. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences*, 14, 1090. <https://doi.org/10.3390/educsci14101090>
- Kim, J., Klopfer, M., Grohs, J. R., Eldardiry, H., Weichert, J., Cox, L. A., & Pike, D. (2025). Examining faculty and student perceptions of generative AI in university courses. *Innovative Higher Education*. Advance online publication.
- Kolade, O., Owoseni, A., & Egbetokun, A. (2024). Is AI changing learning and assessment as we know it? Evidence from a ChatGPT experiment and a conceptual framework. *Heliyon*, 10, e25953. <https://doi.org/10.1016/j.heliyon.2024.e25953>
- Luo, J., Zheng, C., Yin, J., & Teo, H. H. (2025). Design and assessment of AI-based learning tools in higher education: A systematic review. *International Journal of Educational Technology in Higher Education*, 22(1), 42. <https://doi.org/10.1186/s41239-025-00541-0>
- Luo, T. (2024). Academic integrity in the era of generative AI: A comparative policy analysis of global universities. *Journal of Higher Education Policy and Management*, 46, 257–274.
- Mah, D.-K., & Groß, N. (2024). Artificial intelligence in higher education: Exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal*

- of Educational Technology in Higher Education, 21(1), 58.
<https://doi.org/10.1186/s41239-024-00483-6>
- Nikolic, S., Daniel, S., Haque, R., Belkina, M., Hassan, G. M., Grundy, S., & Sandison, C. (2023). ChatGPT versus engineering education assessment: A multidisciplinary and multi-institutional benchmarking and analysis. *European Journal of Engineering Education*, 48, 559–614. <https://doi.org/10.1080/03043797.2023.2213169>
- Sharma, D., & Saxena, A. B. (2023). Exploring the role of artificial intelligence for augmentation of adaptable sustainable education. *Asian Journal of Advanced Research and Reports*, 17(11), 179–184.
- Sharma, D., & Saxena, A. B. (2026). Machine learning-based detection of students at academic risk due to excessive AI tool usage. In *New Horizons of Science, Technology and Culture* (Vol. 11, pp. 12–26).
- Smolansky, A., Cram, A., Radulescu, C., Zeivots, S., Huber, E., & Kizilcec, R. F. (2023). Educator and student perspectives on the impact of generative AI on assessments in higher education. In *Proceedings of the 10th ACM Conference on Learning@Scale* (pp. 378–382). <https://doi.org/10.1145/3573051.3596191>
- Thanh, B. N., Vo, D. T. H., Nhat, M. N., Pham, T. T. T., Trung, H. T., & Xuan, S. H. (2023). Race with the machines: Assessing the capability of generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*, 39, 59–81. <https://doi.org/10.14742/ajet.8902>
- Weng, X., Zhu, M., & Kimmons, R. (2025). Pedagogical applications of generative AI in higher education: Trends, challenges, and future research directions. *TechTrends*. <https://doi.org/10.1007/s11528-025-01100-1>
- Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International Journal of Educational Technology in Higher Education*, 21, 40. <https://doi.org/10.1186/s41239-024-00468-z>