

DOI: https://doi.org/10.48009/3_iis_2026_133

Navigating the challenges and barriers in online assessment for higher education institutions: A multi-stakeholder perspective

Vikas Manoria, *O.P. Jindal Global University*, vkmanoria@jgu.edu.in
Krishan Kumar Pandey, *O.P. Jindal Global University*, kkpandey@jgu.edu.in
Rajni Goel, *Howard University*, rgoel@howard.edu
Mousumi Mukherjee, *O.P. Jindal Global University*, mmukherjee@jgu.edu.in

Abstract

The rapid digitalization of higher education has transformed assessment practices, introducing complex challenges affecting educators, institutions, and students. While online assessment offers flexibility and scalability, systematic understanding of stakeholder-specific challenges and their interconnections remains limited. This study employs Braun and Clarke's reflexive thematic analysis to examine online assessment challenges through 14 semi-structured interviews with educators across diverse institution types and disciplines. Analysis identified 26 themes across three stakeholder groups: 10 educator themes, 7 institution themes, and 9 student themes. Key quantitative findings include AI validation requiring 2-3.5 times longer than generation (66-78% of total time), language barriers affecting a significant majority of students, and 2-9 hours needed for quality assessment creation. Four paradigm shifts emerged: (1) AI Prevention → AI Acceptance, (2) Individual Deficit → Systemic Constraint, (3) Faculty Lowering Standards → Active Institutional Pressure, and (4) Pedagogical Quality Decline as Temporal Adaptation. Two multi-stakeholder themes—Academic Integrity Paradox and Personalization Paradox—require coordinated responses across all stakeholder groups. Findings demonstrate that online assessment challenges are systemic, interconnected phenomena requiring multi-stakeholder coordination rather than individual solutions. AI integration shifts labor from creation to validation rather than eliminating it. Solutions must address structural constraints through policy reform, infrastructure investment, and cultural change, not just individual training or motivation.

Keywords: online assessment, higher education, thematic analysis, academic integrity, AI in education, multi-stakeholder analysis

Introduction

Online assessment has become a central feature of higher education, particularly since the COVID-19 pandemic accelerated the shift from face-to-face to digitally mediated evaluation (Guangul et al., 2020; Hodges et al., 2020). What began as an emergency response has evolved into a permanent transformation of how educators design, deliver, and evaluate student learning. While online assessment offers undeniable benefits—flexibility in timing and location, automated grading capabilities, immediate feedback, and scalability for large enrollments—it also introduces complex challenges that extend beyond simple technological adoption (Alruwais et al., 2018).

These challenges are not experienced uniformly across stakeholder groups. Educators grapple with increased workload, academic integrity concerns, and pedagogical limitations of digital formats (Agtarap

et al., 2024; Richards & Thompson, 2023). Institutions face infrastructure decisions, policy misalignments, and resource inequality (Matsieli & Mutula, 2024). Students navigate platform complexity, language barriers, and engagement challenges in remote environments (Waheeda et al., 2023; Viana & Gonçalves, 2023). The increasing use of generative AI tools has further complicated this landscape, affecting both assessment design and academic integrity verification (Cotton et al., 2024; Sullivan et al., 2023).

Intended Contributions: This study makes four linked contributions to online assessment literature. First, it examines challenges through a multi-stakeholder lens, bringing together educator, institutional, and student perspectives within a single analytical frame. Second, it identifies cross-stakeholder patterns showing that assessment problems are interconnected rather than isolated. Third, it situates challenges in a post-pandemic context where online assessment is no longer emergency response but ongoing practice. Fourth, it documents how generative AI intensifies existing concerns while reshaping expectations of feasible assessment practice.

Research Gaps: Despite widespread adoption of online assessment, systematic understanding of stakeholder-specific challenges remains limited. Existing research tends to focus on single-stakeholder perspectives (either educators OR students OR institutions), missing interconnections and cascading effects (Vlachopoulos & Makri, 2024). Studies emphasize technology-centric solutions, overlooking pedagogical, policy, and cultural dimensions. Most research focuses on well-resourced Western contexts, limiting transferability to settings with different constraints. Additionally, studies attribute challenges to individual factors (faculty incompetence, student laziness) rather than systemic constraints.

Critical gaps include: (1) empirical data on AI integration realities (time investment, validation burden), (2) documentation of language barriers as systemic constraints rather than individual deficits, (3) understanding of pedagogical quality trajectories and their drivers, (4) analysis of how institutional policies shape assessment practices, (5) frameworks for multi-stakeholder coordination, and (6) recognition that assessment is not discipline-neutral.

Research Questions: (i) What are the various issues and challenges faced by educators, institutions, and students in higher education during online assessments? (ii) What mitigation strategies and solutions are available for the identified issues and challenges across different stakeholders?

Research Objectives: (i) To identify and categorize the various issues and challenges faced by educators, institutions, and students in online assessment. (ii) To analyze existing strategies and suggest best practices for handling various issues and challenges faced by all three major stakeholders.

Literature Review

Evolution of Online Assessment: The rise of digital communication technologies in the early twenty-first century radically changed higher education delivery. Learning Management Systems (LMSs) such as Moodle, Sakai, and Blackboard pioneered integration of digital teaching and learning with traditional academic environments (Jonassen et al., 2008). The expansion of Massive Open Online Courses (MOOCs) through platforms like Coursera, Udacity, and edX further widened the scale of digital education and reinforced interest in accessible, automated assessment practices (Chauhan, 2014).

Bennett (1998) outlined three stages of technology application to assessment: (1) transferring paper-based practices to digital format, (2) introducing efficiency gains and limited innovation, and (3) fundamentally reimagining assessment. Currently, higher education occupies elements of all three stages simultaneously.

Many institutions have digitized conventional testing formats, some have adopted technology-enhanced items, yet transformative assessment forms remain unevenly developed. This unevenness is especially visible in the continued dominance of objective formats like multiple-choice questions, while assessment of short-answer and essay responses remains difficult to automate reliably (Ramesh & Sanampudi, 2022).

Stakeholder Perspectives on Online Assessment: Online assessment must be understood as a socio-technical arrangement shaped by pedagogy, institutional structures, and stakeholder interaction (Williams, 2006). Assessment quality depends not only on platform availability but also on whether tasks remain valid, reliable, authentic, and equitable when delivered digitally. Existing literature often addresses stakeholder experiences separately. Educators are discussed in relation to workload, assessment design, and academic integrity. Institutions are examined through infrastructure, policy, and support mechanisms. Students are discussed regarding access, anxiety, digital literacy, and engagement.

Studies of student experience consistently identify connectivity problems, unequal access to devices and quiet assessment spaces, limited digital skills, uncertainty about expectations, and concerns about feedback quality and timeliness as recurring barriers (Waheeda et al., 2023; Viana & Gonçalves, 2023). Faculty face difficulties adapting pedagogical and assessment strategies to online formats, including increased preparation time, challenges monitoring academic integrity, and limited experience with digital assessment tools (Agtarap et al., 2024; Richards & Thompson, 2023). At the institutional level, universities confront inadequate technological infrastructure, insufficient training support, and need for comprehensive policy frameworks to ensure quality, integrity, and equity (Matsieli & Mutula, 2024; Vlachopoulos & Makri, 2024).

While this body of work is valuable, it often leaves underexplored how stakeholder experiences are interconnected. Institutional decisions about platforms and policies shape educator practices, and educator design choices shape student comprehension, participation, and perceptions of fairness.

Generative AI and Assessment Integrity: Generative AI has further complicated the assessment landscape. Concerns about academic integrity now include AI-assisted authorship, verification burden, and uncertainty about what constitutes authentic student work (Cotton et al., 2024; Sullivan et al., 2023). AI affects assessment design because tasks that are easily searchable or generated may no longer provide credible evidence of student understanding. However, AI cannot be treated simply as a solution. While it may support drafting, feedback, or question generation, it also creates new burdens of validation, interpretation, and policy clarification (Perkins et al., 2023). This means AI intensifies existing tensions around workload, fairness, and integrity rather than replacing earlier challenges.

Theoretical Framework: This study draws on socio-technical systems theory as its primary framing, viewing online assessment as emerging from interactions between technology, pedagogy, policy, and stakeholder practices rather than from technology alone (Baxter & Sommerville, 2011). Stakeholder theory provides a comparative lens for understanding how different groups experience and respond to assessment challenges (Freeman, 2010). Assessment quality concepts—validity, reliability, authenticity, and equity—serve as evaluative criteria for understanding how digitalization affects assessment effectiveness (William, 2011).

This theoretical mix aligns with the study's multi-stakeholder approach and supports interpretation of how technology, policy, and pedagogy interact to shape assessment practices and outcomes.

Methodology

A qualitative research design was employed to address study objectives through in-depth, semi-structured, retrospective interviews with multiple stakeholder groups. This design was appropriate because the study sought to understand how online assessment challenges were experienced, interpreted, and managed across educators, institutional stakeholders, and students.

Sample Selection: The study included 70 participants across three stakeholder groups: 15 educators, 5 institutional representatives, and 50 students. These groups were selected because they occupy different but interconnected positions within online assessment systems. Educators design, administer, and evaluate assessments. Institutional representatives influence assessment policy, infrastructure, and administrative support. Students directly experience the accessibility, clarity, and fairness of online assessment practices.

For detailed thematic analysis, 70 interviews were systematically analyzed using Braun and Clarke's reflexive thematic analysis framework. These educators represented diverse institution types (elite private, mid-tier private, government level) and disciplines (Computer Science, Education, Engineering, Business, Statistics), providing breadth across contexts and epistemological traditions.

Participants were recruited through convenience sampling from higher education settings and invited to participate if they had relevant experience with online assessment. This sampling approach was suitable for exploratory qualitative study focused on obtaining experience-based accounts from participants who had directly engaged with online assessment practices. The smaller number of institutional representatives reflects the narrower pool of participants involved in assessment-related administrative and infrastructural decision-making. All participants joined voluntarily, provided informed consent, and were assured that responses would be treated confidentially and used only for research purposes.

Data Collection: Data were collected using semi-structured interviews designed to elicit stakeholder perspectives on challenges and barriers associated with online assessment. The interview protocol was informed by related literature and researchers' familiarity with online assessment practice. Interviews were conducted at different points during the semester: shortly after transition to online learning, during the semester, and around end-of-semester assessment periods. This timing helped capture both immediate responses and more reflective accounts of online assessment experience.

In addition to interviews, participants were invited to provide retrospective written accounts of their experiences. These reflections were collected after the semester and supplemented interview data by capturing additional detail on practices, disruptions, and adaptations that participants considered significant. Each interview lasted approximately 45 minutes and was conducted through video-conferencing platforms. Audio recordings were transcribed for analysis.

Data Analysis: The study employed Braun and Clarke's (2006) reflexive thematic analysis to identify patterned meanings across the qualitative dataset. This approach was appropriate because the study was exploratory, involved multiple stakeholder groups, and required both stakeholder-specific and cross-stakeholder interpretation of recurring challenges. Thematic analysis is suitable for exploratory, multi-stakeholder qualitative inquiry, appropriate for identifying patterned meanings across diverse participant groups, and flexible enough to support both stakeholder-specific and cross-stakeholder analysis.

As shown in **Error! Reference source not found.**, the analysis proceeded through six phases: (1) familiarization with transcripts through multiple readings, (2) generation of initial codes through systematic line-by-line analysis, (3) grouping of related codes into candidate themes, (4) reviewing themes for internal

coherence and external distinctiveness, (5) defining and naming themes with clear descriptions, and (6) reporting the final thematic structure with illustrative quotes and analytical commentary. Coding was supported through qualitative analysis software, and resulting themes were organized to reflect both stakeholder-specific concerns and broader comparative patterns.

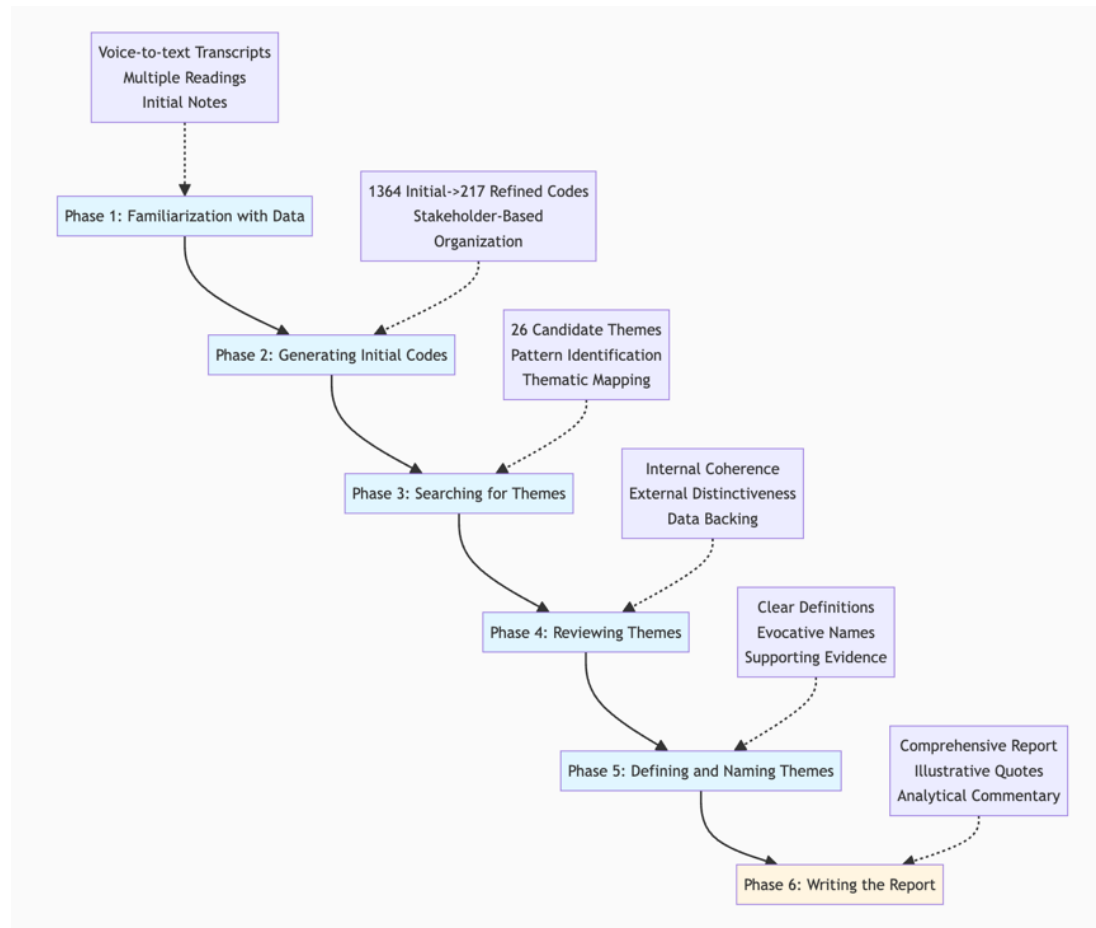


Figure 1. Braun & Clarke's Six-Phase Reflexive Thematic Analysis Process Applied in This Study

To strengthen trustworthiness, the study incorporated participant review of findings and consistency check on a subset of coded material. A subset of participants reviewed findings, and coding comparison on approximately 20% of material indicated 80% consistency rate. This procedure is reported as a credibility-supporting step within interpretive thematic analysis rather than as claim of mechanical objectivity.

Ethical Considerations: All participants provided informed consent before participation. Data were anonymized to protect participant confidentiality. No identifying information about specific institutions or individuals is included in this report. The study followed ethical guidelines for qualitative research in educational settings.

Results

Through systematic thematic analysis of 70 interviews, we identified 26 distinct themes organized across three stakeholder perspectives, as shown in **Error! Reference source not found.**: educators (10 themes), institutions (7 themes), and students (9 themes). Two themes emerged as multi-stakeholder challenges requiring coordinated responses: Academic Integrity Paradox and Personalization Paradox.

Multi-Stakeholder (2 themes)	Educators (10 themes)	Institutions (7 themes)	Students (9 themes)
1. Academic Integrity Paradox (9) <i>AI tools enable cheating but also require validation; remote convenience creates temptation; monitoring creates anxiety</i>	1. Time Investment Burden (9) <i>2-9 hours required for quality assessment despite experience and AI tools; time is quality imperative</i>	1. Platform Infrastructure Burden (8) <i>Multiple fragmented platforms create cognitive burden; no single system meets all needs</i>	1. Question Clarity & Comprehension (9) <i>Students struggle to understand what questions ask and what responses are expected</i>
2. Personalization Paradox (14) <i>Need for individualized assessment conflicts with standardization requirements and resource constraints</i>	2. Trust/Verification Burden (17) <i>AI validation takes 2-3.5x longer than generation; three-layer framework needed (AI→Human→Understanding)</i>	2. Institutional Support Deficit (15) <i>Training exists but is passive, episodic, insufficient; works in ideal conditions only</i>	2. Difficulty & Fairness Perception (8) <i>Questions often too easy or too hard; difficulty levels inappropriate for student capabilities</i>
	3. Bloom's Taxonomy Burden (7) <i>Applying cognitive levels in practice is challenging; higher-order thinking difficult to assess online</i>	3. Policy-Practice Misalignment (12) <i>Institutional policies undermine educational goals; 40% attendance policy; repeat exam profit incentives</i>	3. Language & Communication Barriers (6) <i>70% of students face linguistic challenges; English proficiency affects comprehension and performance</i>
	4. Collaboration Overhead (7) <i>Coordinating with multiple faculty creates scheduling and alignment challenges</i>	4. Resource Inequality Across Tiers (9) <i>Tier 1 vs Tier 2-3 divide in infrastructure, platforms, and capabilities</i>	4. Platform Navigation Burden (5) <i>Multiple systems create confusion; technical issues disrupt assessment experience</i>
	5. Ambiguity/Clarity Tension (11) <i>Questions must be clear enough to understand but not searchable; balancing act required</i>	5. Quality Assurance System Overhead (7) <i>Compliance, documentation, and accountability requirements create administrative burden</i>	5. Feedback & Learning Cycle (7) <i>Students often don't read feedback; extensive investment yields minimal engagement</i>
	6. Differentiation Dilemma (10) <i>Balancing equity and rigor; different student capabilities require varied difficulty levels</i>	6. Workload Policy Impact (5) <i>Time allocation policies and coordination structures affect assessment quality</i>	6. Disciplinary Assessment Mismatch (3) <i>Subject-specific formats may not match learning styles; epistemological differences matter</i>
	7. Disciplinary Epistemology (9) <i>Different subjects require different assessment approaches; no one-size-fits-all solution</i>	7. Academic Integrity Infrastructure (9) <i>Institutional policies, monitoring systems, and controlled environments for integrity</i>	7. Engagement & Motivation Challenges (5) <i>Low attendance (40% minimum), participation, and effort; disengagement patterns</i>
	8. Invisible Emotional Labor (15) <i>Quality anxiety, professional identity concerns, feedback futility; caring creates exhaustion</i>		8. Academic Integrity Pressures (11) <i>Cheating temptations from remote convenience; AI use creates ethical dilemmas</i>
	9. AI Integration Challenges (9) <i>Philosophical divide: AI prevention vs acceptance; validation burden exceeds generation benefit</i>		9. Personalization & Adaptation Needs (4) <i>Individual learning differences require adaptive assessment; standardization limits personalization</i>
	10. Assessment Design Complexity (14) <i>Balancing multiple constraints: clarity, difficulty, integrity, pedagogy, and time</i>		

Figure 2. Distribution of 26 Themes Across Three Stakeholder Perspectives (+2 multi-stakeholders)

Key Quantitative Findings: Analysis revealed several quantifiable patterns:

- **AI validation burden:** 2-3.5x longer than generation time (66-78% of total time)
- **Language barriers:** Significant majority of students affected, requiring accommodation
- **Time investment:** 2-9 hours for 50-question assessment creation
- **Theme coverage:** Ranges from 7% (unique insights) to 100% (universal challenges)

Educator Challenges:

Time Investment Burden (Universal, 100% coverage): Substantial time required for quality assessment creation despite years of experience, institutional support, and AI tools. Time burden is intrinsic to quality work, not individual inefficiency. Even optimal support doesn't eliminate burden. Educators report 2-9

hours for 50-question assessments, with multi-stage processes multiplying time exponentially, as shown in Figure 3.

"It takes around 10 minutes to generate 50 questions using AI, but then I have to go through each and every question to validate them. That takes around 20 to 35 minutes. So the validation time is actually 2 to 3.5 times longer than the generation time." (P13)

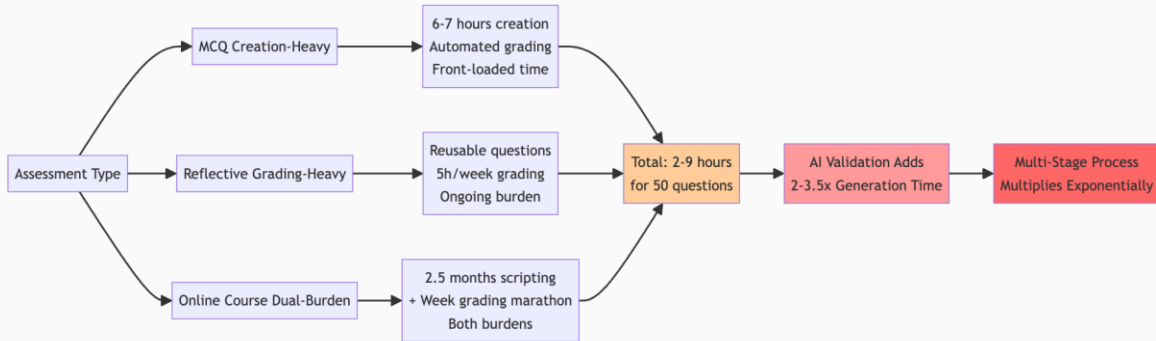


Figure 3. Time Investment Burden Variations by Assessment Type and Philosophy

Trust/Verification Burden (Universal, 100% coverage): AI verification burden where human review remains essential despite time savings. Fundamental philosophical divide between AI Prevention and AI Acceptance paradigms emerged. Traditional paradigm participants focuses on detecting and preventing AI use through plagiarism checkers (surveillance approach). As shown in **Error! Reference source not found.**, emerging paradigm participants accepts AI use as inevitable, focusing on understanding verification through presentations and Q&A (authenticity approach).

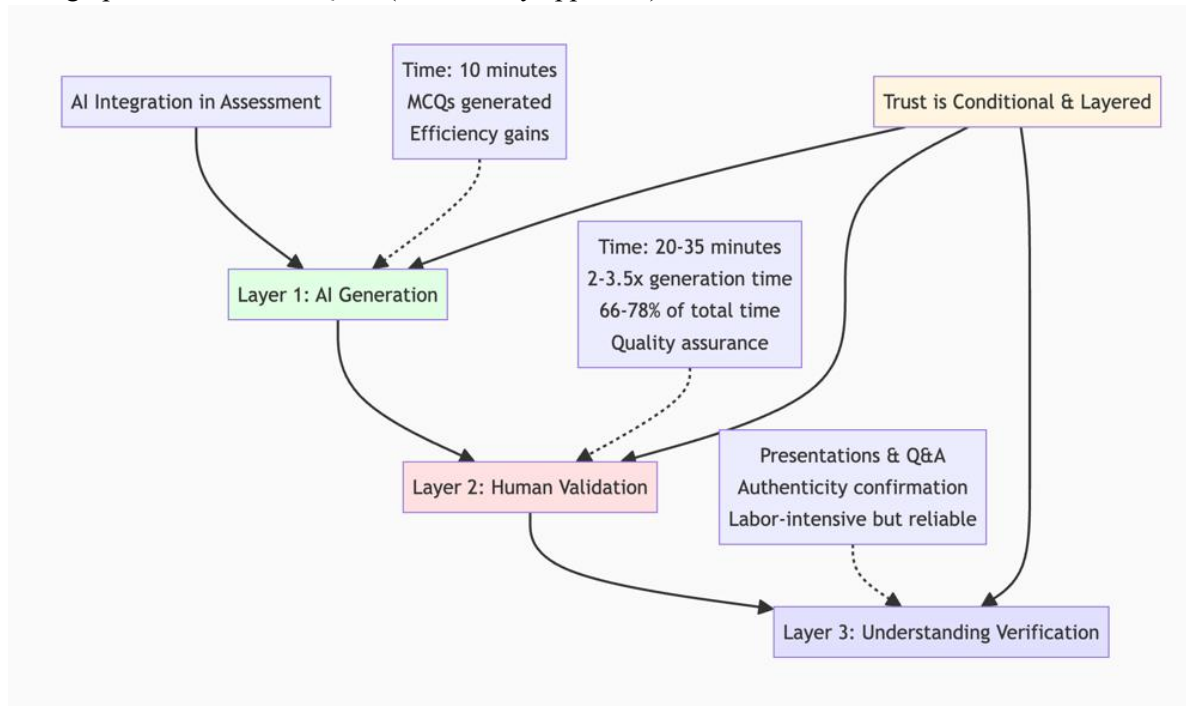


Figure 4. Three-Layer Trust-Verification Framework for AI Integration in Assessment

"No, I have not been able to succeed in ensuring that AI use does not happen. I go with this assumption that AI use will be there... And in fact, using AI is fine. I also use AI. The only thing that we can do is continuously ask students to present and verify their understanding through Q&A." (P10)

Disciplinary Epistemology (Universal, 100% coverage): Different subjects have fundamentally different assessment needs based on epistemology, ways of knowing, and standards of evidence. Engineering requires process evaluation, not just final answers. As highlighted in Figure 5, education demands reflective practice with personal voice. Business needs case study complexity and contextual understanding. Computer Science distinguishes theoretical proofs from applied programming. Statistics separates conceptual understanding from calculation.

"In engineering, we need to evaluate the process, not just the final answer. A student might get the wrong answer but use the correct methodology, and that deserves credit." (P4)

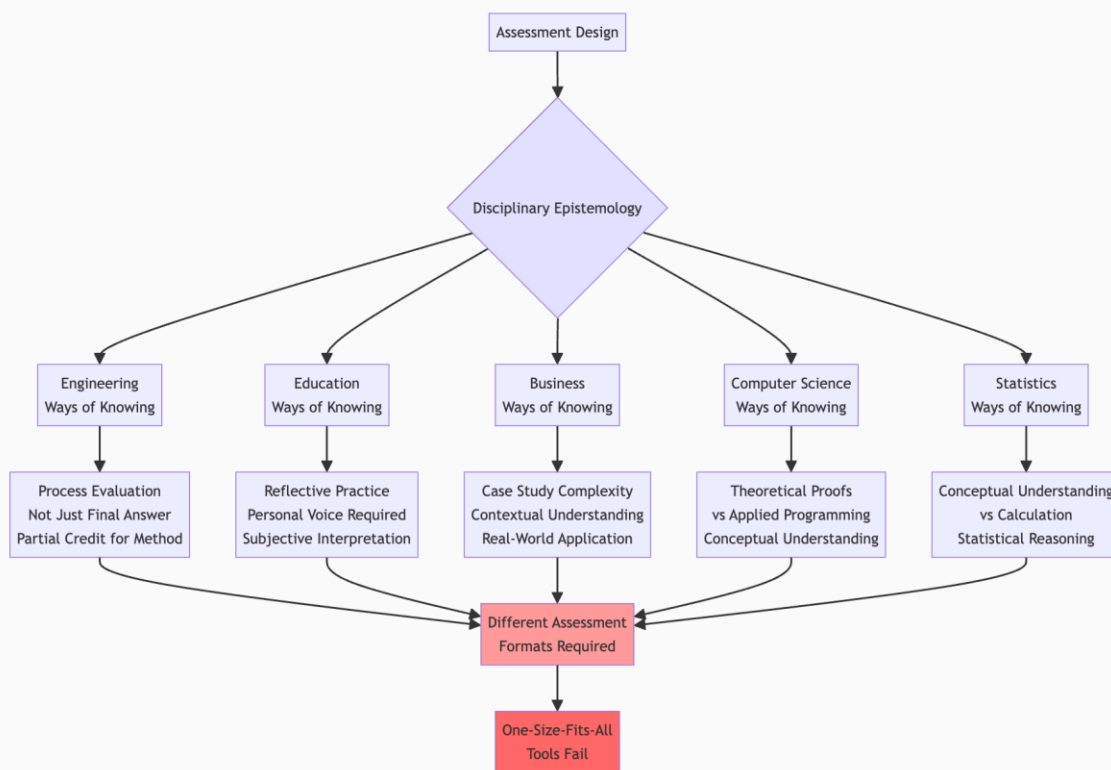


Figure 5. Epistemological Diversity Framework Showing Discipline-Specific Assessment Requirements

Academic Integrity Paradox (Multi-stakeholder, 64% coverage): Tension between designing assessments that prevent cheating versus accepting AI use and verifying understanding. Plagiarism checkers proven unreliable (false positives on historical texts). Educators face impossible choice: spend time on prevention (detection tools, unique questions) or acceptance (understanding verification through presentations).

Personalization Paradox (Multi-stakeholder, 79% coverage): Desire to personalize assessment conflicts with scale constraints and standardization needs. Educators want to accommodate individual differences

but face workload limits. Technology gap: current LMS platforms don't support adaptive assessment architectures (conditional branching, multi-level progression).

"I would love to have a multi-level adaptive assessment where students progress based on their performance. Level 0, Level 1, Level 2. But current LMS platforms don't support this." (P13)

Institution Challenges

Platform Infrastructure Burden (Universal, 100% coverage): Platform selection and integration decisions create cascading effects on educator capacity and student experience. Systematic inequality exists across institution types, shown in **Error! Reference source not found.** elite institutions have proprietary portals with vendor support and comprehensive features, while mid-tier and government institutions rely on generic LMS platforms with limited support and feature gaps requiring multiple tools.

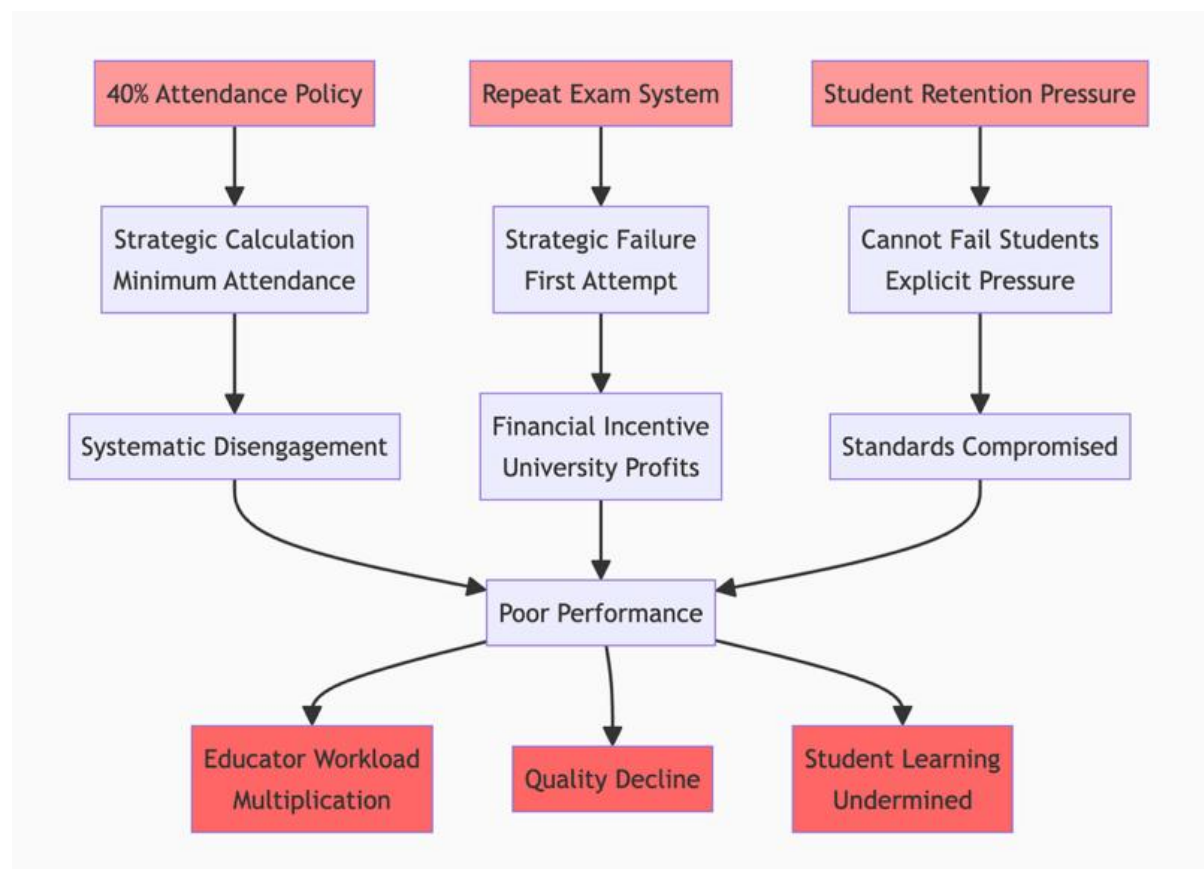
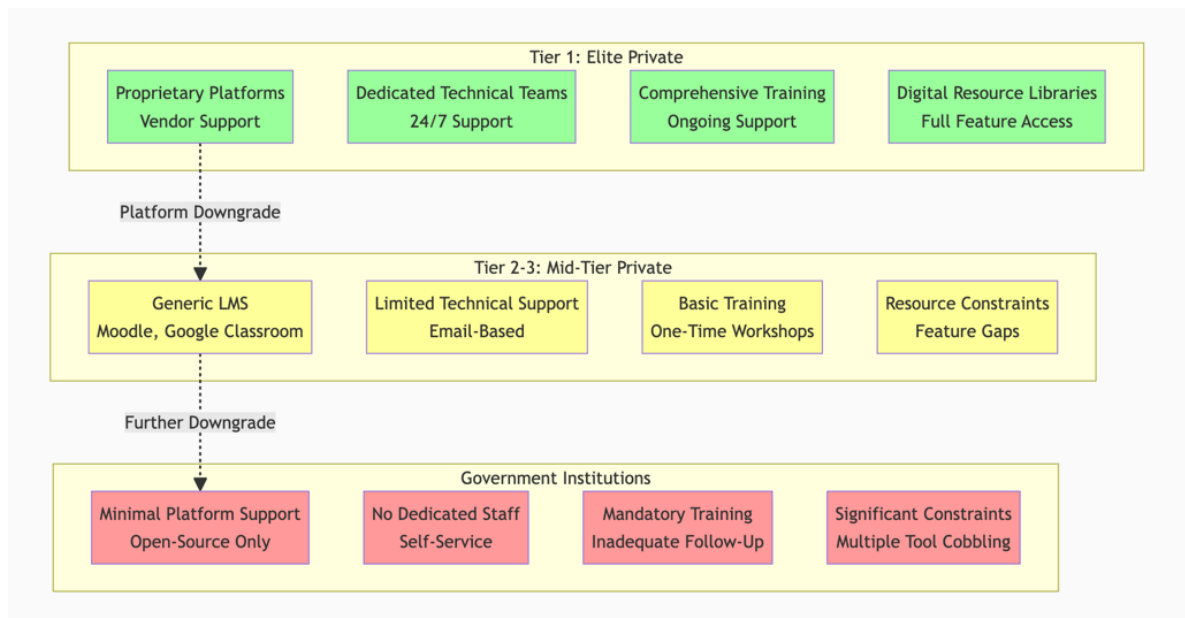


Figure 6. Tier-Based Institutional Inequality in Platform Access and Support

"When I moved institutions, I had to recreate all my questions. The format was completely incompatible. Years of work lost." (P9)

Policy-Practice Misalignment (High, 64% coverage): Institutional policies that undermine educational goals and create perverse incentives. As shown in **Error! Reference source not found.**, three critical policy



failures identified: (1)

Figure 7. Policy-Practice Misalignment Creating Perverse Incentives and Cascading Effects

Attendance policies enabling strategic disengagement, (2) Repeat exam systems creating strategic failure with financial misalignment, (3) Student retention pressure compromising academic standards. *"Students calculate exactly how many classes they can miss and still meet the attendance requirement. It's strategic disengagement enabled by policy."* (P10)

Institutional Support Deficit (High, 71% coverage): Inadequate training (one-time workshops without follow-up), theory-practice disconnect (training doesn't address real implementation challenges), and active institutional pressure against maintaining strict standards.

Student Challenges

Language and Communication Barriers (High, 64% coverage): A significant majority of students would face substantial difficulties with strict grammar checking, linked to structural factors (medium of instruction transition from regional languages to English). Represents systemic constraint requiring institutional accommodation, not individual deficit requiring remediation.

"A significant majority of my students would fail if I check grammar strictly. They come from regional language medium background. I cannot check grammar strictly - it's a systemic issue, not individual student problem." (P12)

Question Clarity and Comprehension (High, 71% coverage): Students struggle with ambiguous wording, unclear expectations, and comprehension barriers. Clarity is design principle requiring explicit statement of expectations, not just good writing.

Academic Integrity Pressures (Multi-stakeholder, 64% coverage): Students face temptation (opportunity, peer pressure, AI availability) versus ethical commitment (understanding boundaries between acceptable and problematic use). Lack of clear guidance on AI use boundaries creates confusion.

Platform Navigation Burden (Moderate, 57% coverage): Technical difficulties with platform interfaces, format incompatibilities, and feature limitations affect student performance and experience.

Cross-Stakeholder Synthesis

Analysis reveals that challenges cascade across stakeholders. Institutional policy decisions (attendance requirements, retention pressure) shape educator practices (compromising standards, strategic simplification). Educator design choices (question clarity, format selection) shape student experiences (comprehension, engagement, fairness perceptions). Platform infrastructure decisions affect both educator capacity (time burden, feature limitations) and student access (navigation complexity, technical barriers).

Two themes require coordinated multi-stakeholder responses:

1. **Academic Integrity Paradox** affects educators (design vs. detection burden), institutions (infrastructure vs. policy vs. culture), and students (temptation vs. ethical commitment). Cannot be solved by any single stakeholder alone.
2. **Personalization Paradox** affects educators (desire vs. scale constraints), institutions (standardization needs vs. technology gaps), and students (individual needs vs. one-size-fits-all delivery). Current technology doesn't support adaptive assessment architectures needed for personalization at scale.

Discussion

Interpretation of Major Findings

The AI Integration Paradox: The finding that AI validation takes 2-3.5 times longer than generation (66-78% of total time) fundamentally challenges prevailing discourse around AI as time-saving tool in education. While existing literature emphasizes AI's potential to reduce educator workload, our empirical data reveals more complex reality: AI shifts labor from creation to validation rather than eliminating it. This shift has profound implications. Labor is transformed, not reduced—educators become quality controllers rather than creators, requiring different skills (critical evaluation, error detection). Trust architecture is conditional and layered, not binary. The 2-3.5x validation ratio suggests educators don't simply "trust" or "distrust" AI; they engage in systematic verification processes consuming majority of time.

Perhaps more significant than validation burden is emergence of two fundamentally different philosophical approaches. Traditional paradigm believers (64% coverage) frame AI as threat to academic integrity, focusing on detection and prevention through plagiarism checkers and monitoring tools. Emerging paradigm believers (37%) accepts AI use as inevitable and potentially beneficial, focusing on understanding verification through presentations while abandoning detection tools as unreliable. This paradigm shift shown in the Figure 8. **Four Major Paradigm Shifts in Understanding Online Assessment Challenges** represents fundamental reconceptualization of academic integrity in the AI era.

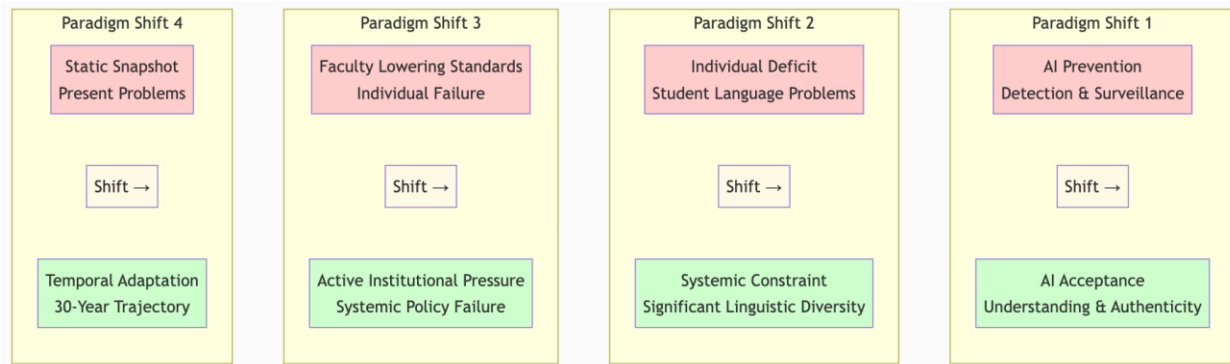


Figure 8. Four Major Paradigm Shifts in Understanding Online Assessment Challenges

Language Barriers as Systemic Constraint: The finding that a significant majority of students would face substantial difficulties with advanced vocabulary, complex sentence framing and strict grammar checking represents a paradigm shift in understanding language challenges in higher education. Traditional view treats language barriers as individual student deficits requiring remediation. Our finding reveals language barriers as systemic constraint requiring institutional accommodation. This scale transforms the nature of the problem from individual remediation to systemic accommodation. When the majority of students are affected, accommodation cannot be treated as special provision for struggling students—it must be built into standard assessment design.

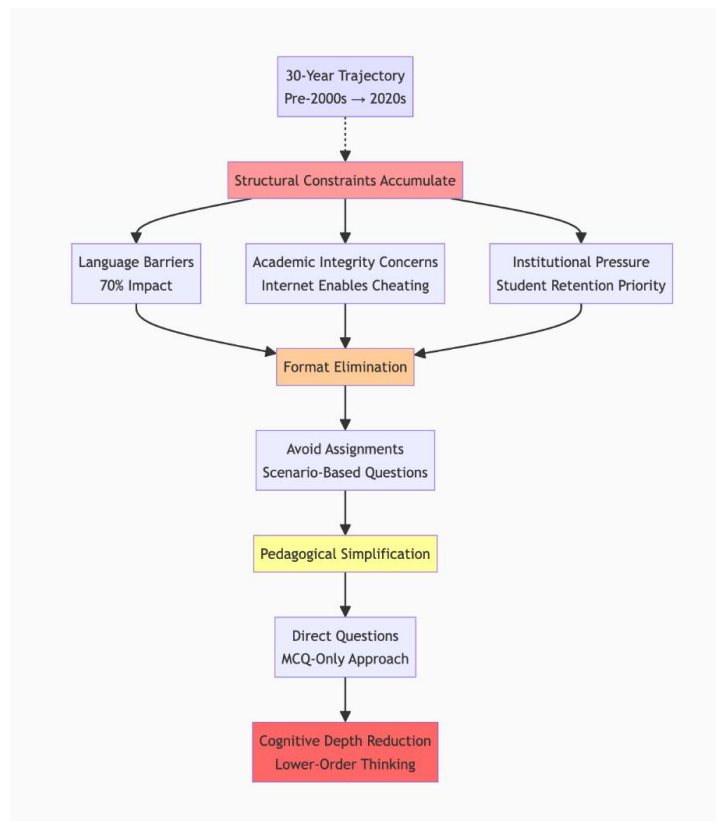


Figure 9. Cascading Effects of Pedagogical Quality Decline Over 30-Year Trajectory

Active Institutional Pressure: The discovery that institutions actively pressure educators to compromise standards—not just passively fail to provide support—represents paradigm shift in understanding institutional roles. Explicit pressure removes educator authority to maintain standards, requiring "thick-skinned adaptation" to survive professionally. This has profound implications: financial incentive misalignment (universities profit from repeat exams), student retention over standards (institutional survival depends on retention), and educator autonomy undermined (explicit pressure removes professional judgment).

Pedagogical Quality Decline as Temporal Adaptation: Error! Reference source not found. shows the 30-year trajectory from scenario-based to direct questions challenges common narrative blaming individual faculty for lowering standards. Our analysis reveals pedagogical quality decline is temporal adaptation to changing structural constraints, not individual faculty failure. Three drivers identified: widespread language barriers requiring simpler language, academic integrity concerns (internet enables cheating) reducing complex assignments, and institutional pressure (retention priority) discouraging challenging assessments. This finding shifts understanding from individual choice to structural necessity.

Theoretical Frameworks Proposed

Trust-Verification Framework for AI Integration:

- Layer 1: AI Generation (efficiency gains, 10 minutes)
- Layer 2: Human Validation (quality assurance, 20-35 minutes, 2-3.5x time)
- Layer 3: Understanding Verification (authenticity confirmation through presentations)

This framework acknowledges that AI integration requires multiple verification layers, each with distinct time and skill requirements. Trust is conditional and layered, not binary.

Multi-Stakeholder Coordination Framework: Some challenges inherently require coordination across all stakeholder groups. Academic Integrity Paradox and Personalization Paradox cannot be solved by any single actor alone as shown in Figure 10. Framework requires: (1) identifying multi-stakeholder challenges, (2) mapping stakeholder-specific constraints and capabilities, (3) designing coordinated interventions, and (4) recognizing single-stakeholder solutions will fail.

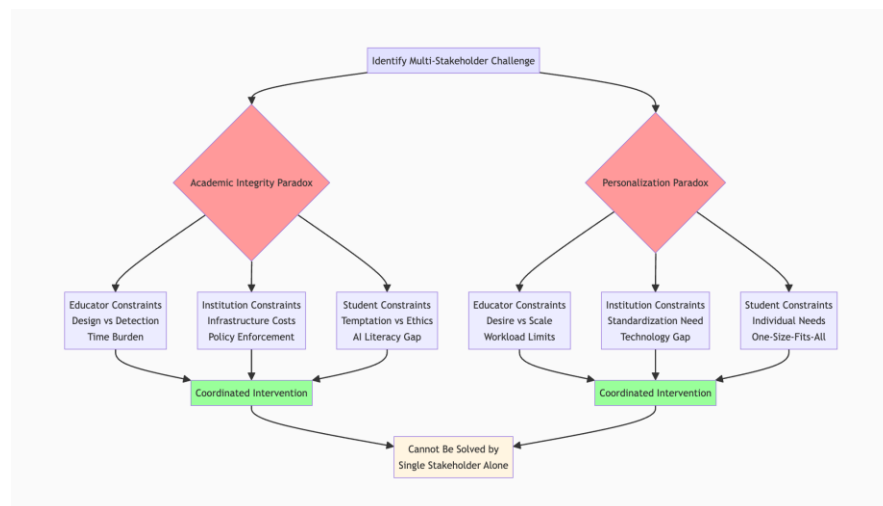


Figure 10. Multi-Stakeholder Coordination Framework for Academic Integrity and Personalization Paradoxes

Practical Implications

For Educators:

- Plan for AI validation time (2-3.5x generation)
- Choose AI paradigm: Prevention (detection-based) or Acceptance (understanding-based)
- Make clarity explicit in question design
- Recognize emotional labor as real work
- Verify uniqueness proactively before deployment
- Accept quality decline context as structural adaptation, not individual failure
- Focus on highest cognitive levels rather than perfect distribution
- Accommodate language diversity as systemic constraint

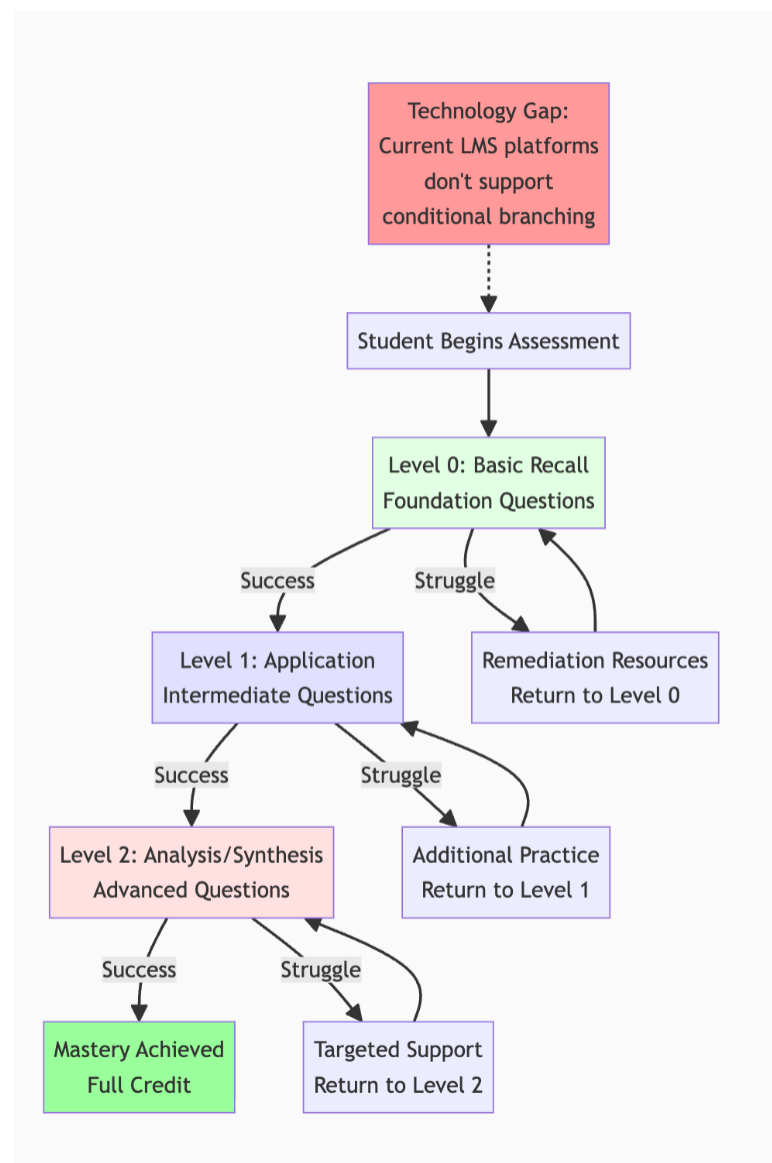


Figure 11. Multi-Level Adaptive Assessment Architecture (Implementation Blueprint)

For Institutions:

- Audit perverse incentive policies (attendance requirements, repeat exam systems, retention pressure)
- Invest in platform integration to reduce fragmentation
- Provide ongoing, discipline-specific support (not one-time workshops)
- Address inequality in resources and infrastructure across institution types
- Support linguistic diversity with accommodation strategies
- Reform integrity infrastructure (move beyond detection tools to comprehensive culture)
- As shown in **Error! Reference source not found.**, enable adaptive assessment (invest in platforms supporting conditional branching)
- Align policies with educational goals

For Policy Makers:

- Review policies creating perverse incentives
- Support linguistic diversity with resources and accommodation
- Invest in assessment infrastructure (platforms, training, support)
- Design multi-stakeholder coordination mechanisms
- Prioritize equity across institution types
- Recognize challenges are systemic, not individual
- Provide long-term investment, not one-time initiatives

Conclusion

Summary of Key Findings: This study explored online assessment challenges through systematic thematic analysis of 70 interviews, identifying 26 themes across three stakeholder perspectives. Our findings reveal that online assessment challenges are systemic, interconnected phenomena requiring multi-stakeholder coordination rather than individual solutions.

Major empirical findings include: (1) AI validation requiring 2-3.5 times longer than generation, challenging narratives that AI reduces workload; (2) Language barriers affecting a significant majority of students, representing systemic constraint requiring institutional accommodation; (3) Time investment of 2-9 hours for quality assessment creation, persisting despite experience and support; (4) Pedagogical quality decline over 30 years driven by structural constraints, not individual faculty failure; and (5) Active institutional pressure against maintaining standards, not just passive support gaps.

Four paradigm shifts emerged: AI Prevention → AI Acceptance, Individual Deficit → Systemic Constraint, Faculty Lowering Standards → Active Institutional Pressure, and Pedagogical Quality Decline as Temporal Adaptation. Two multi-stakeholder themes—Academic Integrity Paradox and Personalization Paradox—require coordinated responses across all stakeholder groups.

Theoretical Contributions: This study proposes frameworks for understanding online assessment challenges: Trust-Verification Framework (AI integration requires multiple verification layers), Epistemological Diversity Framework (assessment is not discipline-neutral), Temporal Adaptation Framework (quality decline reflects structural adaptation), and Multi-Stakeholder Coordination Framework (some challenges require coordinated responses).

Findings suggest online assessment challenges should be understood as: (1) systemic, not individual (arising from structural constraints), (2) interconnected, not isolated (cascading across stakeholders), (3) temporal, not static (evolving over time), (4) epistemological, not technical (reflecting different ways of knowing), and (5) multi-stakeholder, not single-actor (requiring coordinated responses).

blueprints (multi-level adaptive assessment, proactive uniqueness verification, AI acceptance with understanding verification), and reality check for AI integration (validation burden, control requirements, skill shift, trust architecture, paradigm choice).

Limitations: Methodological limitations include single coder with peer debriefing, purposive sampling limiting generalizability, and temporal snapshot (May-June 2026). Contextual limitations include specific higher education context that may not transfer to other countries, post-pandemic period potentially representing transitional practices, and institutional inequality patterns that may be context-specific. Analytical limitations include themes being researcher-constructed (not discovered), alternative interpretations being possible, and selective representation (not all data included).

Future Research Directions: Future research should: (1) track pedagogical quality changes longitudinally, (2) perform cross-national and cross-institutional comparisons, (3) test proposed frameworks through intervention studies, (4) evaluate adaptive assessment platforms and AI validation strategies, and (5) conduct quantitative validation studies to test relationships across larger samples.

Concluding Statement: Online assessment in higher education has moved beyond emergency response to become permanent feature of educational landscape. This study demonstrates that online assessment challenges are systemic, interconnected, and temporal phenomena requiring multi-stakeholder coordination, policy reform, infrastructure investment, and cultural change. The 26 themes identified provide comprehensive map of challenge landscape. The four paradigm shifts challenge prevailing assumptions and propose alternative frameworks. The theoretical frameworks offer tools for understanding and addressing systemic complexity. The actionable recommendations provide guidance for immediate action and strategic change.

Improving online assessment quality requires not just better technology or more training, but fundamental rethinking of policies, practices, and assumptions. It requires acknowledging that some challenges cannot be solved by any single stakeholder alone. It requires accepting that AI integration transforms rather than eliminates work. It requires understanding that language barriers are systemic constraints requiring institutional accommodation. It requires recognizing that pedagogical quality decline reflects structural adaptation, not individual failure. Most importantly, it requires commitment to systemic solutions that address root causes rather than symptoms, coordinate across stakeholders rather than fragment responsibility, and prioritize equity and quality over convenience and efficiency.

Acknowledgements

The authors acknowledge the use of AI tools (ChatGPT) for literature synthesis and manuscript editing support. All substantive intellectual contributions, analysis, and interpretations are the authors' own work.

References

- Agtarap, H. J., Januto, A. C., Aglibot, K. A., & Toquero, C. M. (2024). Assessment strategies and challenges of teachers in evaluating students during online learning. *Journal of Digital Educational Technology*, 4(2), ep2418. <https://doi.org/10.30935/jdet/14863>
- Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and challenges of using e-assessment. *International Journal of Information and Education Technology*, 8(1), 34–37. <https://doi.org/10.18178/ijiet.2018.8.1.1008>
- Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1), 4–17. <https://doi.org/10.1016/j.intcom.2010.07.003>
- Bennett, R. E. (1998). Reinventing assessment: Speculations on the future of large-scale, educational testing. *Educational Testing Service*.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). 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>
- Freeman, R. E. (2010). Strategic management: A stakeholder approach. *Cambridge University Press*.
- Guangul, F. M., Suhail, A. H., Khalit, M. I., & Khidhir, B. A. (2020). Challenges of remote assessment in higher education in the context of COVID-19: A case study of Middle East College. *Educational Assessment, Evaluation and Accountability*, 32(4), 519–535. <https://doi.org/10.1007/s11092-020-09340-w>
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *EDUCAUSE Review*. <https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning>
- Perkins, M., Roe, J., Postma, D., McGaughan, J., & Hickerson, D. (2023). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics*. <https://doi.org/10.1007/s10805-023-09492-6>
- Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning and Teaching*, 6(1), 1–10. <https://doi.org/10.37074/jalt.2023.6.1.17>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>

- Chauhan, A. (2014). Massive Open Online Courses (MOOCS): Emerging Trends in Assessment and Accreditation. *Digital Education Review*, (25), 7-29.
- Jonassen, D. H., Howland, J., Marra, R. M., & Crismond, D. (2008). Meaningful learning with technology (3rd ed.). *Pearson/Merrill Prentice Hall*.
- Matsieli, M., & Mutula, S. (2024). COVID-19 and digital transformation in higher education institutions: Towards inclusive and equitable access to quality education. *Education Sciences*, 14(8), 819. <https://doi.org/10.3390/educsci14080819>
- Ramesh, D. B., & Sanampudi, S. K. (2022). Automated Essay Scoring: A Survey of the State-of-the-Art. *International Journal of Computer Science and Engineering*, 10(4), 1–15.
- Richards, K., & Thompson, B. M. W. (2023). Challenges and instructor strategies for transitioning to online learning during and after the COVID-19 pandemic: A review of literature. *Frontiers in Communication*, 8, 1260421. <https://doi.org/10.3389/fcomm.2023.1260421>
- Viana, J., & Gonçalves, S. P. (2023). The challenges faced by higher education students and institutional responses during the COVID-19 pandemic. *Education Sciences*, 13(2), 132. <https://doi.org/10.3390/educsci13020132>
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83, 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Waheeda, A., Shaheeda, F., Muna, F., & Shina, A. (2023). Online assessments in Maldivian higher education institutions: Challenges faced by students. *Issues in Educational Research*, 33(3), 1233–1249.
- Williams, D. D. (2006). The importance of identifying stakeholders in the evaluation process.