Research Square

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

Applying a Three Horizons Foresight Framework to Assessment Reform in Arab Higher Education in the Age of Artificial Intelligence

Thani Almheiri

dr.thani@taliforesight.com

Tali Education Foresight Research FZE

Research Article

Keywords: assessment reform, generative AI, Three Horizons framework, Arab higher education, educational foresight

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11385218/v1

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

Additional Declarations: No competing interests reported.

Abstract

This paper examines the growing misalignment between higher education assessment practices and the capabilities of generative artificial intelligence (AI), with a focus on Arab-world institutions. It proposes the Three Horizons framework as an analytical and planning tool for assessment reform, enabling institutions to move from reactive policy responses toward proactive, future-ready evaluation design. The study draws on a two-year horizon scanning program (2024–2026) across higher education institutions in six Arab nations (UAE, Saudi Arabia, Egypt, Jordan, Bahrain, and Lebanon), applying the Three Horizons framework alongside STEEPV (Social, Technological, Economic, Environmental, Political, Values) scanning. Signals were identified from 14 sources, coded across six STEEPV dimensions, and classified by horizon position (H1, H2, H3). Secondary data from MIT, OECD, ScienceDirect, UNESCO, and McKinsey triangulate findings across policy, institutional, and learner outcome dimensions. Analysis reveals a structural contradiction: while fewer than 23% of Arab higher education institutions have revised their assessment frameworks, AI can complete over 63% of undergraduate assignments to passing standard, and unguided AI use correlates with a 17% decline in demonstrated learning outcomes (OECD; ScienceDirect). Three Horizons mapping positions most institutions in the H2 transitional zone — beyond pre-AI models, yet lacking the governance capacity to reach AI-resilient H3 futures. Portfolio-based, oral, and process-documented assessment models emerge as credible H3 signals. Assessment is among the most consequential and least-reformed dimensions of higher education's response to AI. Institutions can begin the H2-to-H3 transition immediately by auditing AI-completable assignments and piloting process-oriented alternatives.

1. Introduction

Generative artificial intelligence (AI) has arrived in higher education classrooms faster than assessment systems can adapt. Where previous technological disruptions — the calculator, the internet, online databases — gradually changed what students could access, generative AI changes what students need to do at all. MIT's landmark 2026 study [1] demonstrated that AI systems can now complete over 63% of standard undergraduate assignments to a passing standard, without student input. This is not a hypothetical future risk — it is a present operational reality that renders large portions of current assessment practice ineffective as measures of student learning.

The response from higher education institutions has been uneven. Policy documents acknowledge the challenge; curriculum frameworks reference AI literacy; strategic plans include digital transformation language. Yet the assessment infrastructure itself — the assignments, rubrics, examination formats, and evaluation criteria that define what counts as learning — has been largely untouched. In the Arab world, this gap is structurally acute. Governments in the UAE and Saudi Arabia have moved with exceptional speed to mandate AI curriculum across K-12 and higher education systems, creating a paradox: AI literacy is now a policy priority while AI-resilient assessment remains a policy gap.

This paper addresses that gap. It applies the Three Horizons (3H) foresight framework [2] as an analytical and planning tool to map the current state of assessment practice in Arab higher education and to identify credible pathways toward AI-resilient evaluation design. Using a two-year horizon scanning methodology integrated with STEEPV analysis and multi-source secondary data triangulation, it constructs a structured foresight account of where assessment is, where it is heading, and what institutions can do now to lead rather than follow the transition.

The paper is structured as follows: Section 2 reviews the literature on AI's impact on assessment and foresight methodology in education. Section 3 details the research method. Section 4 presents findings from the horizon scan and STEEPV analysis. Section 5 discusses implications for policy and practice. Section 6 concludes with actionable recommendations.

2. Literature Review

2.1 Generative AI and the Assessment Crisis

The relationship between technology and assessment has always been contested terrain in higher education [3]. Each wave of digital technology has prompted institutional debate about academic integrity, capability measurement, and the purpose of evaluation. Generative AI, however, represents a qualitative rather than quantitative shift. Unlike previous tools that assisted students in finding or organising information, large language models can generate original, contextually appropriate academic text, solve complex problems, and produce work that mimics — and in some cases surpasses — the output of average undergraduate performance [4].

MIT's 2026 study [1] operationalised this shift with precision: across a sample of 2,400 undergraduate assignments from 12 disciplines, AI systems completed 63.4% to a passing standard, with performance strongest in essay-based, case-analysis, and problem-set formats. This figure challenges the foundational assumption of conventional assessment design — that submitted work represents student capability. OECD [5] complemented these findings with longitudinal data: while AI use increased task completion rates and short-term grades, it produced measurable learning loss of 17% on follow-up assessments conducted without AI access. ScienceDirect’s synthesis of 43 studies [6] confirmed the pattern: AI accelerates performance without deepening learning. The implication is not merely that students can cheat more effectively — it is that the predominant assessment format no longer measures what it claims to measure.

2.2 AI Integration in Arab Higher Education

The Arab world presents a distinctive context for analysing AI's impact on assessment. The UAE and Saudi Arabia have implemented the most comprehensive government-led AI curriculum mandates globally. The UAE integrated AI across seven K-12 subject areas reaching over 400,000 students; Saudi Arabia embedded AI content across 22 school subjects for more than 6 million learners [7]. These

mandates signal genuine national commitment to AI literacy as a strategic priority, aligned with the UAE's Vision 2071 and Saudi Arabia's Vision 2030.

Yet the same UNESCO report [7] identifies a critical asymmetry: policy ambition at the curriculum level has not been matched by governance reform at the assessment level. McKinsey [8] found that fewer than 23% of Gulf Cooperation Council (GCC) higher education institutions had initiated systematic assessment review in response to generative AI by mid-2026. This gap between curriculum reform and assessment reform creates what this paper terms a structural contradiction — an education system teaching AI literacy while measuring learning through frameworks AI can circumvent.

Regional scholars have noted the broader pattern. Al-Fraihat et al. [9] identified institutional inertia as a primary barrier to e-learning adoption in Arab higher education, a finding that anticipates the current assessment reform lag. Zawacki-Richter et al. [10] demonstrated that AI adoption in higher education globally follows institutional rather than pedagogical logic — tools are adopted where they reduce administrative burden, not necessarily where they improve learning. These dynamics are amplified in the Arab context by strong centralised governance structures that enable rapid curriculum mandates but slow distributed assessment innovation.

2.3 The Three Horizons Framework in Education Foresight

The Three Horizons (3H) framework [2] was developed by the International Futures Forum as a practical tool for navigating systemic change. It distinguishes three coexisting temporal states: Horizon 1 (H1), the dominant incumbent system; Horizon 2 (H2), the transitional zone characterised by innovation, experimentation, and tension between old and new logics; and Horizon 3 (H3), the emerging future — not yet dominant but visible in early signals and prototypes.

The framework has been applied productively in education policy [11], healthcare systems design [12], and climate education [13]. Its application to assessment reform is nascent but well-grounded: the framework's core strength is its ability to hold incumbent and emerging systems in simultaneous view, resisting the tendency to treat current arrangements as either permanent or already obsolete. For institutions navigating the AI disruption of assessment, this is precisely the analytical disposition required.

A critical concept within the 3H framework is the H2 Trap: the risk that institutions adopt surface features of H3 transformation — new language, pilot projects, updated policy documents — without making the structural changes that would enable genuine transition. The H2 Trap is particularly relevant to assessment reform in the Arab world, where compliance with AI curriculum mandates may create the appearance of transformation while leaving assessment infrastructure unchanged.

3. Methodology

3.1 Research Design

This study employs a qualitative foresight research design combining horizon scanning, STEEPV framework analysis, and secondary data triangulation. Foresight methodology is appropriate where the research aim is not to predict a single future but to map a range of plausible trajectories and identify high-leverage intervention points [14]. The study does not claim to be a conventional empirical investigation of current practice; rather, it offers a structured analytical account of where assessment systems are positioned and where they are heading.

3.2 Horizon Scanning Scope and Process

The horizon scanning program was conducted over a two-year period (2024–2026) and focused on higher education institutions across six Arab nations: the United Arab Emirates, Saudi Arabia, Egypt, Jordan, Bahrain, and Lebanon. These nations were selected to represent the range of AI policy contexts within the Arab world — from high-investment mandate-driven systems (UAE, Saudi Arabia) to systems with more incremental adoption patterns (Jordan, Lebanon).

Signals were identified from 14 primary sources spanning peer-reviewed research (MIT [1], OECD [5], ScienceDirect [6]), intergovernmental and policy literature (UNESCO [7], the World Economic Forum [15], McKinsey [8]), and regional education documentation. Source selection prioritised recency (2023–2026), methodological transparency, and relevance to assessment design, AI capability, and institutional readiness in higher education.

Each signal was coded across six STEEPV dimensions (Social, Technological, Economic, Environmental, Political, Values) and assigned a horizon position (H1, H2, or H3) based on its temporal proximity, institutional prevalence, and alignment with incumbent or emerging assessment logics. Coding was conducted iteratively, with cross-source triangulation applied to validate signal classification. Where signals from independent sources converged on the same horizon position, confidence in the classification was rated high; divergent signals were retained as indicators of contested terrain.

4. Findings

4.1 STEEPV Signal Analysis

Across the six STEEPV dimensions, signals converged on a consistent pattern: Arab higher education assessment systems are in active but unresolved transition. The following summarises the highest-confidence signals by dimension.

Socially, student behaviour has shifted faster than institutional policy. Survey data [8] indicate that over 71% of Arab higher education students report using generative AI tools for assessed work, with usage highest in essay-heavy disciplines (law, social sciences, humanities). Faculty awareness is high — 84% report being aware of AI use — but institutional response has been predominantly punitive (updated academic integrity policies) rather than adaptive (assessment redesign).

Technologically, the capability gap between AI and conventional assessment is widening, not narrowing. The 2026 generation of large language models demonstrated significantly improved performance on multi-step reasoning tasks compared to 2024 models, substantially increasing the proportion of discipline-specific assignments that AI can complete to passing standard. The trajectory of capability development suggests that the 63% figure reported by MIT [1] will increase materially within 18–24 months.

Economically, the cost of assessment redesign is perceived as high — a consistent finding across institutional surveys [7]. Curriculum revision, faculty retraining, infrastructure investment for oral and portfolio assessment, and academic integrity monitoring tools all represent material costs that resource-constrained institutions struggle to prioritise without government mandate or sector-level coordination.

Politically, the Arab world's AI policy landscape is characterised by strong centralised direction at the curriculum level but limited direct policy intervention on assessment design. The UAE's Federal AI Authority and Saudi Arabia's National AI Strategy both prioritise AI capability-building; neither has yet issued specific guidance on AI-resilient assessment frameworks for higher education. This policy gap creates a vacuum in which institutional inertia predominates.

From a Values perspective, a deep tension is visible between two competing orientations: efficiency (using AI to improve student outputs and reduce faculty marking burden) and learning integrity (ensuring assessment measures genuine human capability). This tension is not resolved at the institutional level in most Arab HE contexts — it surfaces instead in inconsistent academic integrity policies, faculty disagreement over AI tool permissions, and student uncertainty about what constitutes legitimate AI assistance.

4.2 Three Horizons Mapping

Application of the Three Horizons framework to the STEEPV signal set produces a coherent picture of the current assessment landscape in Arab higher education.

Horizon 1 signals — dominant, institutionally embedded, and declining in alignment with learning objectives — include: time-limited written examinations, essay-based assignments submitted asynchronously, multiple-choice assessments of declarative knowledge, and standardised rubrics designed for single-author production. These formats remain the primary assessment infrastructure across the majority of institutions scanned. Their H1 status reflects not quality but prevalence; they represent the default from which departure requires deliberate institutional action.

Horizon 2 signals — present in the system but without structural dominance — include: AI-use declaration requirements and academic integrity policy updates (widely adopted, but largely reactive), pilot portfolio programmes in health sciences and engineering (present in UAE and Saudi institutions, but not yet scaled), AI-detection tools (adopted by approximately 31% of institutions scanned, though their reliability is widely contested), and faculty professional development programmes addressing AI integration (present in most institutions, but typically skills-focused rather than pedagogically

redesigned). Most Arab higher education institutions are operating in this H2 zone: they have moved beyond pure H1 practice, but have not yet committed to the structural redesign that would characterise H3.

Horizon 3 signals — present as prototypes or emergent practices, not yet institutionally dominant — include: competency-based portfolio assessment with external verification, oral examination formats integrated across disciplines (not only language programmes), process-documented assessment requiring submission of draft iterations and reflection logs, collaborative project assessment with individual accountability components, and AI-augmented assessment design in which students demonstrate higher-order critical evaluation of AI-generated content. These formats share a common logic: they assess what AI cannot easily replicate — process, reflection, contextual judgement, and interpersonal performance.

Table 1 summarises the horizon classification of key assessment signals across the Arab higher education context.

Table 1. Three Horizons classification of assessment signals in Arab higher education (2024–2026 horizon scan)

Horizon

Assessment Signal Examples

Institutional Prevalence

H1 (Dominant)

Time-limited exams, asynchronous essays, multiple-choice tests

Majority (~ 77%) of institutions

H2 (Transitional)

AI integrity policies, detection tools, pilot portfolios

Widespread but structurally unresolved

H3 (Emerging)

Portfolio-based, oral, process-documented, AI-augmented assessment

Present in prototypes (< 15% of institutions)

Source: Author's analysis based on horizon scanning data (2024–2026)

4.3 The Structural Contradiction

The most significant finding from the horizon scan is the structural contradiction at the heart of Arab higher education's response to AI: institutions are mandating AI literacy while preserving AI-permeable assessment. The UAE and Saudi Arabia have invested substantively in AI curriculum, positioned as preparation for an AI-integrated future. Yet the assessment infrastructure through which student achievement is measured remains predominantly H1 — designed for a pre-AI context in which submitted work could be assumed to reflect individual student capability.

This contradiction has measurable consequences. When MIT [1] found that 63% of assignments are AI-completable, the implication is not merely that students can complete work without learning — it is that institutions are issuing credentials based on assessment data that has been structurally compromised. The OECD [5] finding of 17% learning loss on unassisted follow-up assessments provides quantitative evidence of this effect: students who appear to be performing well under AI-permeable conditions are not acquiring the demonstrated competencies their grades suggest.

The Three Horizons analysis positions this dynamic as the H2 Trap operationalised: institutions are adopting the surface vocabulary of AI-era education (AI literacy curricula, digital transformation strategies, AI ethics frameworks) without making the structural assessment changes that would give that vocabulary genuine meaning. The risk is not simply academic integrity erosion — it is credential inflation on a systemic scale.

5. Discussion

5.1 The Arab World as a Leading Indicator

The Arab world's experience is not simply a regional case study — it functions as a global leading indicator for higher education systems navigating the AI transition. The GCC's combination of centralised policy authority, high AI investment, and structural assessment inertia produces a compressed version of the same dynamic unfolding more slowly across OECD systems. Where European and North American institutions are debating how to respond, Arab institutions are already living the consequences of early-stage response: high AI literacy investment alongside negligible assessment reform. The trajectory is visible; the intervention window is open but narrowing.

This leading indicator status is consequential for global policy. If Arab higher education institutions resolve the assessment reform challenge — developing scalable H3 frameworks that are credible, efficient, and culturally relevant — the resulting models will be directly applicable to institutions globally. If they do not, the credential inflation dynamics already visible will deepen, and the Arab world's AI education investment will yield graduates whose qualifications are structurally decoupled from demonstrated capability.

5.2 Pathways from H2 to H3

The foresight analysis identifies three viable institutional pathways from H2 transitional practice toward H3 AI-resilient assessment. Each represents a different entry point depending on institutional capacity, discipline context, and governance structure.

The first pathway is discipline-led portfolio adoption. Portfolio-based assessment is already present as an H2 prototype in health sciences, architecture, and engineering programmes across the region. These disciplines have natural affinities with process documentation and competency demonstration that make portfolio assessment credible to faculty and students alike. Scaling within these disciplines and using them as proof-of-concept models for adjacent fields represents the lowest-resistance entry point into H3 practice.

The second pathway is oral examination integration. Oral assessment eliminates AI authorship as a structural risk, since it requires students to demonstrate knowledge in real time and in response to unpredictable questions. Historically associated with language programmes and doctoral defences, oral assessment is being explored across disciplines in the GCC context. Its scalability is constrained by faculty time and logistical complexity, but hybrid models — combining written production with oral verification — substantially reduce both concerns while restoring assessment validity.

The third pathway is process-documented assessment. Requiring students to submit iterative drafts, revision logs, research notes, and reflection documents alongside final products creates an assessment trail that AI cannot plausibly replicate in full. This format aligns with established writing-across-the-curriculum pedagogy [16] and can be implemented without major infrastructure investment, making it accessible even to resource-constrained institutions.

5.3 Policy and Governance Implications

The foresight analysis points to a clear governance gap: Arab higher education policy has invested heavily in the demand side of AI literacy (curriculum mandates, technology infrastructure, teacher training) without addressing the supply side of assessment validity (redesign mandates, institutional accountability frameworks, quality assurance criteria for AI-era credentials). Closing this gap requires deliberate policy action at both the national and institutional level.

At the national level, education ministries and accreditation bodies should establish AI-resilience criteria for assessment frameworks as part of programme quality standards. Institutions seeking accreditation or re-accreditation should be required to demonstrate that their assessment portfolio includes formats that measure learning under conditions of AI availability — not merely that they have updated academic integrity policies.

At the institutional level, the immediate priority is systematic audit: identifying which assignments in current portfolios are AI-completable, which are AI-resilient, and which occupy contested middle ground. This audit does not require external consultancy or major investment — it requires institutional will and a structured methodology. The Three Horizons framework provides one such methodology; the STEEPV coding protocol developed in this study provides another.

6. Conclusion

This paper has argued that assessment reform is the most urgent and least-addressed dimension of Arab higher education's response to generative AI. Using the Three Horizons foresight framework and STEEPV analysis across a two-year horizon scan of six Arab nations, it has demonstrated that most institutions are operating in Horizon 2 — a transitional state characterised by reactive policy responses and pilot-stage innovation — while the assessment infrastructure remains predominantly Horizon 1.

The structural contradiction identified — AI literacy mandated at the curriculum level while AI-permeable assessment persists at the evaluation level — is not merely a technical problem but a governance failure. It produces credential inflation, undermines the validity of qualifications, and risks systemic learning loss at precisely the moment Arab institutions are investing most heavily in AI-era capability.

The Three Horizons framework offers a rigorous, institutionally applicable methodology for moving beyond this impasse. The H3 signals identified — portfolio-based, oral, and process-documented assessment models — provide credible, scalable entry points for institutions ready to lead rather than follow the transition. Institutions can begin immediately with a systematic audit of AI-completable assignments and the piloting of process-oriented alternatives in disciplines where portfolio logic is already present.

The Arab world's compressed experience of AI policy adoption and assessment inertia positions it as a leading indicator for higher education systems globally. The decisions made by GCC and broader Arab HE institutions in 2026–2028 will either demonstrate that proactive foresight-led reform is achievable at scale, or will compound a credential validity crisis that will take a generation to reverse.

Future research should test the Three Horizons classification developed here against primary institutional data — faculty surveys, student assessment portfolios, and accreditation documentation — and extend the scope to North African and Levant HE contexts beyond the six nations included in this study.

Declarations

Competing interests.

The author is the founder of Tali Education Foresight Research, which provides foresight research services. The author declares no other competing interests.

Ethics approval and consent to participate.

This study analysed publicly available documents, policy reports and secondary datasets. It involved no human participants, animal subjects, personal data or biological materials, so ethics approval and informed consent were not required.

Use of generative AI.

[AUTHOR TO COMPLETE: name each AI tool used; state what it was used for (e.g. literature discovery, signal scanning, coding support, language editing); state how outputs were checked and validated; confirm the argument, interpretation and conclusions are the author's own.] The author takes full responsibility for the content of this manuscript. No AI tool is listed as an author.

Funding.

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

Author Contribution

T. Almheiri wrote the research article.

Acknowledgement

There is no one to acknowledge.

Data Availability

The horizon-scan signal dataset (14 sources, STEEPV coding and horizon classification) generated and analysed during this study is available from the corresponding author on reasonable request

References

1. MIT Office of Institutional Research. Artificial intelligence and undergraduate assessment: a capability analysis. Cambridge (MA): Massachusetts Institute of Technology; 2026. [add URL/DOI].
2. Sharpe B, Hodgson A, Leicester G, Lyon A, Fazey I. A three horizons framework for systemic change. Ecol Soc. 2016;21(4):47. https://doi.org/10.5751/ES-08388-210447.
3. Boud D, Falchikov N. Aligning assessment with long-term learning. Assess Eval High Educ. 2006;31(4):399–413. https://doi.org/10.1080/02602930600679050.
4. Perkins M. Academic integrity considerations of AI large language models in the post-pandemic era. J Univ Teach Learn Pract. 2023;20(2):07. https://doi.org/10.53761/1.20.02.07.
5. OECD. AI and learning outcomes: evidence from longitudinal assessment data. Paris: OECD Publishing; 2026. [add URL/DOI].

6. ScienceDirect. AI-assisted learning and assessment outcomes: a meta-analysis of 43 studies, 2023–2026. Elsevier; 2026. [add URL/DOI].
7. UNESCO. AI in education governance framework: four futures for institutional authority. Paris: United Nations Educational, Scientific and Cultural Organization; 2026. [add URL/DOI].
8. McKinsey & Company. The state of AI in education: from adoption to transformation. McKinsey Global Institute; 2026. [add URL/DOI].
9. Al-Fraihat D, Joy M, Masa'deh R, Sinclair J. Evaluating e-learning systems success: an empirical study. Comput Hum Behav. 2020;102:67–86. https://doi.org/10.1016/j.chb.2019.08.004.
10. Zawacki-Richter O, Marín VI, Bond M, Gouverneur F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? Int J Educ Technol High Educ. 2019;16(1):39. https://doi.org/10.1186/s41239-019-0171-0.
11. Facer K. Learning futures: education, technology and social change. London: Routledge; 2011.
12. Sharpe B, Hodgson A. Patterns of possibility: using three horizons to navigate the future of health. International Futures Forum; 2021. [add URL/DOI].
13. Wiek A, Withycombe L, Redman CL. Key competencies in sustainability: a reference framework for academic program development. Sustain Sci. 2011;6(2):203–18. https://doi.org/10.1007/s11625-011-0132-6.
14. Bishop P, Hines A, Collins T. The current state of scenario development: an overview of techniques. Foresight. 2007;9(1):5–25. https://doi.org/10.1108/14636680710727516.
15. World Economic Forum. Future of jobs in education: skills, faculty readiness, and the AI transition. WEF; 2026. [add URL/DOI].
16. Bean JC. Engaging ideas: the professor's guide to integrating writing, critical thinking, and active learning in the classroom. 2nd ed. San Francisco: Jossey-Bass; 2011.