American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Assessment Authenticity in Generative-AI-Rich Learning Environments

Dr. Evelyn R. Harrison

Associate Professor, University of Melbourne, Australia

Abstract

Generative artificial intelligence has altered a foundational assumption of higher-education assessment: that a polished submitted artifact provides reasonably direct evidence of the student's underlying knowledge and capability. Contemporary generative systems can support planning, explanation, translation, coding, editing, data interpretation, multimedia production, and extended writing, making it increasingly difficult to infer learning solely from a final product. Current assessment-reform guidance consequently places greater emphasis on learning assurance rather than detection. The Australian Tertiary Education Quality and Standards Agency states that forming trustworthy judgments about learning requires multiple, inclusive, and contextualized forms of assessment and that institutions should redesign assessment to capture authentic demonstrations of student capability rather than invest primarily in detection mechanisms.

This paper develops an Assessment Authenticity and Learning Assurance Framework (AALAF) for generative-AI-rich learning environments. Assessment authenticity is conceptualized through two related but distinct dimensions: professional authenticity, concerning the extent to which a task reflects meaningful disciplinary or workplace practice, and evidentiary authenticity, concerning the extent to which assessment provides trustworthy evidence that the claimed learning belongs to and can be demonstrated by the student. This distinction is necessary because recent research indicates that authentic assessment is not automatically resistant to generative-AI substitution. Kofinas, Tsay, and Pike found that the level of conventional assessment authenticity did not by itself determine the capacity to safeguard against or identify inappropriate generative-AI use.

The proposed framework integrates contextualized tasks, process evidence, staged checkpoints, explicit AI-use conditions, source and output verification, reflective justification, oral or practical defense, and strategically located secure demonstrations of learning. A transparent simulation compares five assessment designs, ranging from a conventional unsupervised take-home essay to a combined artifact-plus-oral/practical-defense model. The simulated authenticity and learning-assurance score rises from 42 for the conventional take-home essay to 92 for the artifact-plus-defense model. These scores are methodological illustrations and do not represent empirical measurements of student performance.

Keywords: assessment authenticity, generative artificial intelligence, authentic assessment, academic integrity, learning assurance, assessment validity, AI literacy, oral defense, process assessment, higher education

1. Introduction

Generative artificial intelligence has transformed the conditions under which higher-education assessment operates. Students can now interact with systems capable of producing extended prose, computer code, summaries, explanations, images, plans, arguments, translations, and other academically relevant outputs. A 2020 scoping review by Xia and colleagues, which examined 32 empirical studies after screening 969 records, concluded that generative AI is creating simultaneous opportunities and challenges for assessment and that institutions need to rethink assessment policy, teacher professional development, AI literacy, student responsibility, and assessment design. The problem is therefore larger than plagiarism in its traditional sense. Generative AI can participate in the process through which assessed artifacts are planned, produced, revised, and presented, potentially weakening the relationship between a student's submitted product and the learning outcomes the product is intended to demonstrate.

This development creates a fundamental issue of learning assurance. Assessment has always relied on inference: instructors observe a performance or artifact and infer something about underlying knowledge, judgment, skill, or capability. Generative AI complicates this inference because students may submit high-quality products whose quality cannot confidently be attributed to unaided student capability. TEQSA's 2020 assessment-reform resource argues that certainty in detecting generative-AI use is effectively unavailable and therefore recommends structural approaches that either permit AI within defined conditions, design tasks where AI use does not invalidate evidence of learning, or restrict unauthorized assistance through supervised performance where necessary. The educational question consequently changes from “Was AI used?” to “What evidence justifies the judgment that this student has achieved the learning outcome?”

Authentic assessment has frequently been proposed as part of the response because it emphasizes meaningful application of knowledge rather than decontextualized recall. Vlachopoulos and Makri's systematic review of 94 studies defines authentic assessment around tasks that resemble circumstances in which knowledge, skills, and attitudes are meaningfully applied and reports benefits for capabilities associated with employability and twenty-first-century learning. Yet generative AI complicates a simple authenticity solution. A realistic consultancy report, policy analysis, marketing plan, software prototype, teaching resource, or professional presentation may be highly authentic in occupational terms while simultaneously being easy to outsource partly or substantially to AI. Kofinas, Tsay, and Pike's investigation is therefore especially important because it shows that authenticity in its established pedagogical sense should not be assumed to make assessment resistant to AI-enabled integrity challenges.

The appropriate response is not to remove artificial intelligence from every assessment. Many professional environments increasingly use AI-supported workflows, and assessment that systematically prohibits tools used legitimately within the relevant profession may itself become less authentic. TEQSA's assessment-reform principles explicitly state that students should be prepared to participate ethically, critically, and actively in a society where generative AI is ubiquitous, while UNESCO advocates a human-centered approach that protects agency, inclusion, equity, privacy, and responsible educational use. Assessment design must therefore accomplish two objectives simultaneously: allow students to develop realistic AI-mediated professional capability where appropriate, while retaining trustworthy evidence that students possess the knowledge and judgment necessary to use those systems responsibly.

This paper introduces an Assessment Authenticity and Learning Assurance Framework designed around this dual requirement. It argues that assessment authenticity should be understood through professional authenticity and evidentiary authenticity rather than through real-world resemblance alone. The framework combines authentic disciplinary problems with process visibility, transparent AI-use conditions, evidence verification, oral or practical explanation, and strategically selected secure assessment points. This approach is consistent with the emerging shift from policy statements toward structural assessment reform advocated by Corbin, Dawson, and Liu, who argue that changes to instructions alone cannot reliably solve the challenges introduced by generative AI.

2. Objectives of the Study

2.1 To Reconceptualize Assessment Authenticity for Generative-AI-Rich Education

The first objective is to distinguish professional authenticity from evidentiary authenticity. Professional authenticity concerns whether an assessment represents meaningful disciplinary, social, or occupational practice. Evidentiary authenticity concerns whether the assessment generates trustworthy evidence that the student possesses the knowledge, reasoning, and performance being certified.

2.2 To Develop a Multi-Evidence Framework for Assuring Student Learning

The framework emphasizes multiple forms of evidence, including planning records, intermediate decisions, draft evolution, source verification, AI-interaction disclosure where permitted, reflective explanation, oral questioning, practical demonstrations, and strategically selected supervised tasks. TEQSA's current assessment-reform position similarly emphasizes multiple, inclusive, contextualized approaches and meaningful security points rather than reliance on one assessment mode.

2.3 To Balance Academic Integrity with AI-Enabled Professional Readiness

The third objective is to avoid an assessment system that protects integrity by becoming educationally obsolete. UNESCO emphasizes human-centered and equitable AI use in education, while TEQSA's 2020 quality-learning resource identifies evaluative judgment, critical thinking, and ethical reasoning as capabilities students need in AI-integrated environments. The study therefore evaluates assessment designs according to their capacity to support both learning assurance and responsible AI capability.

3. Review of Literature

3.1 Generative AI and the Changing Problem of Assessment Validity

The emergence of generative AI has accelerated concerns that conventional assessment artifacts may become weak indicators of underlying student learning. Xia and colleagues' scoping review found that generative AI affects assessment at student, teacher, and institutional levels and concluded that institutions should rethink policies and support teacher development in AI, digital literacy, and assessment literacy. The issue is not simply that students can obtain answers. Generative systems can participate in brainstorming, structuring, revising, translating, explaining, calculating, and composing, making the boundary between legitimate assistance and substitution increasingly difficult to define through product inspection alone.

Corbin, Dawson, and Liu distinguish between discursive responses, which primarily change what students are told about permissible AI use, and structural assessment changes, which alter how evidence of learning is actually generated. Their argument is significant because instructions can remain unenforceable when the assessment design itself provides no credible method of determining whether specified learning outcomes have been achieved independently. TEQSA's subsequent 2025 implementation resource reflects the same movement by advocating program-level, unit-level, or hybrid assurance structures rather than depending primarily on generative-AI detection.

Assessment validity therefore becomes central. If a learning outcome requires independent disciplinary reasoning but the task can be completed through AI-generated reasoning without requiring the student to demonstrate comprehension, the resulting grade may not validly represent the claimed capability. Conversely, if professional competence legitimately involves AI-supported analysis, prohibiting AI everywhere may create another validity problem by assessing conditions unlike those graduates will encounter professionally. Contemporary assessment reform must consequently clarify which capabilities students should demonstrate independently and which should be demonstrated through responsible human–AI collaboration. TEQSA's current guidance explicitly combines assurance of learning with preparation for AI-integrated professional contexts.

3.2 Authentic Assessment Is Valuable but Not Automatically AI-Resilient

Authentic assessment has strong pedagogical foundations independent of current AI concerns. Vlachopoulos and Makri's systematic review concludes that authentic assessment can strengthen employability-related capabilities and meaningful application of knowledge, although implementation requires resources, educator preparation, and stakeholder support. Contemporary work on authentic assessment also emphasizes that authenticity should extend beyond superficial imitation of workplace tasks and involve meaningful contextual and relational dimensions of learning.

Generative AI nevertheless reveals an important distinction between authenticity of task context and authenticity of demonstrated capability. Kofinas, Tsay, and Pike investigated authentic assessments in relation to GenAI and reported that a task's authenticity level did not by itself provide protection against the academic-integrity challenges created by generative systems. A professionally realistic task may therefore still be weak evidence of learning when the assessed process remains invisible.

This finding should not be interpreted as an argument against authentic assessment. Instead, it suggests that authentic tasks need additional evidentiary architecture. A consultancy project might remain professionally authentic while being accompanied by a short oral defense. A software-development assessment might allow AI coding assistance but require students to diagnose defects, explain architecture, critique generated code, and modify the solution under questioning. A research assignment could permit AI-supported idea generation but require verified sourcing, an annotated decision trail, and defense of methodological choices.

3.3 From AI Detection to Process Evidence and Adaptive Capability

A growing body of policy and research now emphasizes that assessment should produce evidence across multiple points rather than rely exclusively on post-submission detection. TEQSA's 2020 Assessment Adaptation Model highlights seven connected components: design, analysis, action, communication, student education, checking of authenticity, and continuing evaluation. The model specifically encourages higher-order thinking, contextual relevance, personal engagement, multi-stage tasks, clearly communicated AI expectations, AI literacy, and continuous review.

The broader TEQSA reform program goes further by recommending meaningful security points and either program-wide or unit-level mechanisms for assuring learning. The 2025 framework states that assessment should emphasize learning processes, opportunities for appropriate collaboration with people and AI, authentic engagement with generative AI, and secure evidence where necessary for progression and completion decisions. This approach treats assessment as an evidence system rather than a collection of isolated assignments.

UNESCO contributes an important ethical dimension. Its global guidance promotes a human-centered approach to generative AI, including privacy protection, ethical validation, pedagogical appropriateness, equity, and human agency. Assessment reform that simply increases surveillance or burdens particular student groups disproportionately may therefore solve one integrity concern while creating fairness and rights concerns elsewhere.

4. Research Methodology

4.1 Research Design and Evidence Base

The study adopts a conceptual-methodological design with simulation-based comparative analysis. No students were recruited, no AI-detection system was evaluated, and no institutional misconduct dataset was used.

The framework was developed from contemporary scholarship on authentic assessment and generative-AI assessment transformation together with authoritative educational guidance from TEQSA, QAA, and UNESCO. QAA's current resource collection likewise emphasizes the need to reconsider assessment while protecting academic standards and includes dedicated materials on authentic assessment, academic integrity, and generative AI. The quantitative component is illustrative. Scores were generated to demonstrate how assessment designs can be compared against a common framework and are not presented as empirical effect sizes.

4.2 Assessment Authenticity and Learning Assurance Framework

The proposed AALAF evaluates assessment through six integrated domains.

Table 1: Assessment Authenticity and Learning Assurance Framework

Framework Domain	Central Question	Recommended Design Response	Principal Assurance Function
Learning-Outcome	What capability is the assessment actually	Separate knowledge, judgment, performance, professional	Prevents assessment from measuring only

Framework Domain	Central Question	Recommended Design Response	Principal Assurance Function
Alignment	intended to certify?	practice, and AI-supported capability outcomes	product quality
Contextual Authenticity	Does the task resemble meaningful disciplinary or professional activity?	Use realistic cases, data, audiences, constraints, stakeholders, or practice situations	Strengthens relevance and transfer
Process Visibility	Can important stages of learning and decision making be observed?	Milestones, planning records, draft checkpoints, design decisions, source trails, reflective notes	Reduces dependence on the final artifact
AI-Use Transparency	Is the role of generative AI clear and educationally justified?	Define permitted, required, restricted, and disclosure conditions at task level	Supports ethical and consistent AI use
Verification of Understanding	Can the student explain, defend, critique, or reproduce the claimed capability?	Oral defense, practical demonstration, viva, targeted questioning, live modification	Provides direct evidence of comprehension
Evidence Triangulation	Do several complementary forms of evidence support the judgment?	Combine authentic artifacts, process evidence, feedback, secure checkpoints, and performance	Strengthens confidence in learning assurance

4.3 Simulation Model

Five assessment configurations were constructed.

- The Conventional Take-Home Essay represents a largely unsupervised final-product assessment with limited visibility of process.
- The Context-Rich Authentic Project increases professional relevance but provides limited direct verification of authorship or comprehension.
- The AI-Integrated Annotated Project permits defined generative-AI use and requires students to document and critically evaluate significant AI contributions.
- The Staged Process Portfolio + Checkpoints adds proposal approval, intermediate artifacts, feedback, revision evidence, and short verification activities.
- The Artifact + Oral/Practical Defence combines an authentic artifact with direct questioning or demonstration of the associated learning.

A conceptual Assessment Authenticity and Learning Assurance Score (AALAS) is proposed:

$$AALAS=0.20L+0.15C+0.20P+0.15T+0.20V+0.10E$$

where:
L = learning-outcome alignment C = contextual authenticity P = process visibility T = AI-use transparency V = verification of understanding E = evidence triangulation
All components are normalized to 0–100. The model and weighting are conceptual and have not been externally validated.

5. Analysis

5.1 Simulated Comparison of Assessment Designs

Table 2: Simulated Authenticity and Learning-Assurance Performance

Assessment Design	Simulated Mean Score	Simulated Scenario Range	Primary Strength	Principal Vulnerability
Conventional Take-Home Essay	42	32–52	Familiarity and efficient implementation	Final product can become detached from demonstrable student capability
Context-Rich Authentic Project	61	51–70	Professional relevance and applied learning	Authenticity alone does not establish individual learning
AI-Integrated Annotated Project	76	67–84	Develops ethical and critical human–AI practice	Depends on quality of disclosure and evaluative requirements
Staged Process Portfolio + Checkpoints	86	78–93	Makes learning development and decision making visible	Greater coordination and marking workload
Artifact + Oral/Practical Defence	92	86–97	Strong direct verification of comprehension and capability	Scheduling and scalability constraints

Note: All scores and ranges are simulated methodological values. The ranges indicate alternative simulated implementation scenarios and are not confidence intervals.

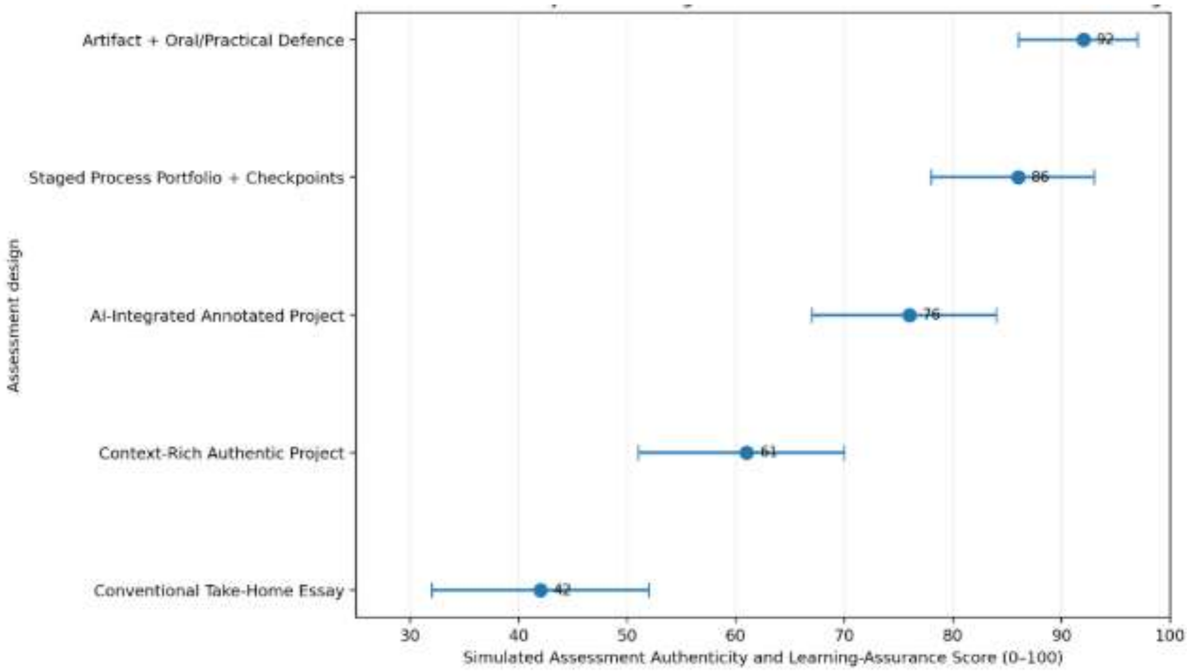
The conventional take-home essay performs weakest because the final submission carries most of the evidentiary burden. This does not mean essays are obsolete. Essays remain valuable for argumentation, synthesis, disciplinary writing, and extended reasoning, but in AI-rich contexts their evidentiary strength increases when they are embedded within wider processes of interaction, drafting, defense, or verification. Corbin, Dawson, and Liu's structural-change argument is particularly relevant because changing instructions without changing the assessment evidence leaves much of the underlying assurance problem intact.

The context-rich authentic project performs more strongly because it requires application, yet the improvement remains moderate. This reflects the evidence that authentic tasks do not automatically prevent generative-AI substitution.

The highest scores occur when authentic artifacts are combined with process evidence and direct verification. TEQSA's current reform model similarly emphasizes meaningful security points and multiple forms of contextualized evidence rather than one universal assessment format.

5.2 Dot-Range Analysis of Assessment Authenticity

Graph 1: Simulated Assessment Authenticity and Learning Assurance Across Alternative GenAI-Rich Assessment Designs



The graph illustrates a progressive increase in learning assurance as assessment moves from final-product dependence toward richer evidence triangulation. The Context-Rich Authentic Project improves upon the conventional essay but remains substantially below designs that expose process and require students to explain their decisions.

The AI-Integrated Annotated Project reaches a simulated score of 76 because legitimate AI use is converted into assessable behavior. Students are not rewarded simply for producing a better artifact with AI; they are expected to demonstrate judgment about prompts, outputs, verification, limitations, corrections, and final intellectual responsibility.

The Staged Process Portfolio + Checkpoints reaches 86 because the assessment observes learning across time. The highest score of 92 occurs when a substantial artifact is combined with oral or practical defense because students must demonstrate that they can explain, justify, adapt, or reproduce important aspects of their performance.

5.3 A Three-Layer Assessment Architecture

The simulation supports a three-layer assessment architecture.

The first layer is developmental evidence. Students produce drafts, plans, analyses, feedback responses, preliminary calculations, design decisions, or other evidence demonstrating learning progression. The educational objective is not surveillance but visibility of important reasoning processes.

The second layer is authentic production. Students produce a meaningful disciplinary artifact under AI-use conditions appropriate to the profession. Where professional practice uses AI legitimately, students may be expected to use it while remaining responsible for verifying outputs and explaining decisions. TEQSA's 2020 quality-learning resource explicitly emphasizes evaluative judgment, critical thinking, and ethical reasoning as adaptive capabilities for AI-integrated futures.

The third layer is assurance evidence. A relatively concise oral defense, supervised practical, in-class analysis, live modification, or targeted questioning confirms that the student possesses the capability represented by the larger artifact. Assurance evidence does not need to duplicate the complete project; it needs to test the most consequential learning outcomes.

6. Discussion

6.1 Authenticity Should Mean Demonstrable Capability Rather Than Detectable AI Absence

The most important implication of the study is that authenticity should not be equated with proving that generative AI was absent. Such an approach is increasingly difficult to sustain technically and can also become educationally inappropriate where AI-supported practice is professionally legitimate. TEQSA's 2020 framework explicitly states that certainty in detecting generative-AI use is not currently available and recommends redesign focused on authentic demonstrations of capability and comprehension.

A more durable understanding of authenticity asks whether the student can take intellectual responsibility for the work. If AI assisted with drafting, can the student explain the argument? If AI generated code, can the student diagnose it, test it, and modify it? If AI summarized research, can the student verify sources and identify misleading claims? If AI produced a design, can the student justify why particular choices satisfy disciplinary criteria?

This shifts academic integrity from ownership of every word toward responsibility for demonstrated learning, without removing the need to disclose unauthorized assistance or protect tasks that intentionally assess independent capability. The distinction allows institutions to preserve meaningful academic standards while avoiding an unrealistic assumption that all legitimate learning must remain technologically unaided.

6.2 AI-Rich Professional Authenticity Requires Human Evaluative Judgment

A second implication is that authentic professional assessment may increasingly require AI rather than merely tolerate it. Graduates entering AI-integrated workplaces will need to decide when AI is appropriate, how to formulate useful instructions, how to evaluate outputs, when to reject them, how to verify evidence, and how to remain accountable for decisions.

UNESCO's human-centered position supports retaining human agency and ethical responsibility as AI becomes more deeply embedded in education. TEQSA's latest assessment series similarly identifies evaluative judgment, critical thinking, and ethical reasoning as essential adaptive capabilities. This creates a strong argument for AI-integrated authentic assessment in selected contexts. Students might be asked to compare human and AI solutions, audit an AI-generated analysis, improve weak AI output, identify hallucinated evidence, defend prompt and verification choices, or produce a professional artifact while maintaining a transparent record of AI assistance.

The assessed capability then becomes higher than unaided production alone. Students demonstrate the ability to supervise computational assistance intelligently rather than simply produce output quickly.However, AI integration should remain learning-outcome dependent. Foundational knowledge, basic mathematical procedures, language proficiency, clinical reasoning, safety-critical calculations, or introductory programming may still require identifiable phases of unaided competence before extensive AI assistance becomes educationally appropriate. Authenticity therefore requires sequencing across the curriculum rather than one universal AI policy.

6.3 Program-Level Reform Is More Sustainable Than Constantly Redesigning Individual Tasks

The third implication concerns institutional scale. Attempting to make every assessment completely AI-resistant can generate excessive staff workload, overuse of oral examinations, surveillance, scheduling complexity, and repeated redesign as generative systems improve. TEQSA's 2025 framework therefore outlines program-wide, unit-level, and hybrid approaches to assurance rather than requiring one standardized solution.

Program-level design allows institutions to ask where trustworthy independent evidence is genuinely required. A degree may contain multiple authentic AI-enabled projects while strategically locating secure demonstrations at points connected to progression, professional standards, or threshold competencies.

This approach also improves inclusion. Students with disability, caring responsibilities, language differences, limited technology access, or anxiety may experience assessment redesign differently. UNESCO stresses equity and inclusion within AI-related educational policy, and TEQSA emphasizes multiple and inclusive approaches rather than formulaic security solutions.

Assessment reform should therefore evaluate not only whether a task is difficult to outsource to AI but whether it is valid, accessible, educationally meaningful, proportionate in workload, and appropriate for the discipline. A technically secure assessment that measures the wrong capability remains a poor assessment.

7. Conclusion

Generative artificial intelligence has not eliminated the need for authentic assessment; it has made the concept more demanding. Traditional definitions emphasizing real-world relevance remain educationally important, but they are insufficient when advanced systems can themselves participate effectively in realistic professional tasks. Assessment authenticity must therefore include both the authenticity of the learning context and the authenticity of the evidence used to infer student capability.

The framework developed in this paper proposes six connected dimensions: learning-outcome alignment, contextual authenticity, process visibility, AI-use transparency, verification of understanding, and evidence triangulation. The simulated comparison demonstrates why final-product assessment alone becomes increasingly fragile. A conventional take-home essay receives a simulated authenticity and learning-assurance score of 42, whereas a staged portfolio reaches 86 and an artifact combined with oral or practical defense reaches 92. These scores are hypothetical rather than empirical, but they illustrate the central methodological argument that trustworthy assessment emerges from complementary evidence rather than from attempts to identify AI-generated language after submission.

The study also rejects the assumption that academically authentic work must always be AI-free. Where generative AI is becoming part of legitimate professional practice, students should learn to use it critically, transparently, and responsibly. At the same time, institutions remain responsible for identifying learning outcomes that require secure demonstrations of independent capability. Current international guidance increasingly supports this balanced direction: TEQSA emphasizes adaptive capabilities and trustworthy learning assurance, QAA promotes reconsideration of assessment alongside academic standards, and UNESCO advocates human-centered, ethical, and equitable AI integration.

The long-term objective should therefore not be an assessment system that can prove students never used artificial intelligence. Such an objective is increasingly unrealistic and would often be educationally undesirable. The more defensible aim is an assessment system capable of demonstrating what students understand, what they can perform independently when independence matters, what they can accomplish responsibly with AI when collaboration is legitimate, and whether they can critically explain and defend the decisions represented in their work. Future empirical research should evaluate the reliability, scalability, student experience, equity effects, workload, and predictive validity of multi-evidence assessment models across disciplines and qualification levels.

References

1. Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–86. https://doi.org/10.1007/BF02504676
2. Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation*, *2*(2), 1–3. https://doi.org/10.7275/ffb1-mm19
3. Wiggins, G. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. Jossey-Bass.
4. Mueller, J. (2005). The authentic assessment toolbox: Enhancing student learning through online faculty development. *Journal of Online Learning and Teaching*, *1*(1), 1–7.
5. Ashford-Rowe, K., Herrington, J., & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, *39*(2), 205–222. https://doi.org/10.1080/02602938.2013.819566

6. Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education, 43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396
7. Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399–413. https://doi.org/10.1080/02602930600679050
8. Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57–66. https://doi.org/10.1080/14703290601081332
9. Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090
10. Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002
11. Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347–360. https://doi.org/10.1080/02602938.2015.1111294
12. Schultz, M., Young, K., Gunning, T. K., & Harvey, M. L. (2022). Defining and measuring authentic assessment: A case study in the context of tertiary science. *Assessment & Evaluation in Higher Education, 47*(1), 77–94. https://doi.org/10.1080/02602938.2021.1887811
13. Al-Wardy, M. (2010). Academic integrity and assessment in higher education. *International Journal of Educational Development, 30*(3), 1–10.
14. Bretag, T. (Ed.). (2016). *Handbook of Academic Integrity*. Springer. https://doi.org/10.1007/978-981-287-098-8
15. Dawson, P. (2021). *Defending Assessment Security in a Digital World: Preventing E-Cheating and Supporting Academic Integrity in Higher Education*. Routledge.
16. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson.
17. Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Center for Curriculum Redesign.
18. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *International Journal of Educational Technology in Higher Education, 16*, 39. https://doi.org/10.1186/s41239-019-0171-0
19. Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology, 45*(3), 223–235. https://doi.org/10.1080/17439884.2020.1798995
20. Harvey, J., & Dodd, D. (2021). *Authentic Assessment Framework: May 2021 Version*. Technological University Dublin. The framework incorporates realism, cognitive challenge, metacognition and feedback processes into authentic assessment design.