Digital-Pedagogy.eu Institute for Education

Journal of Digital Pedagogy 2026, Vol. 5, No. 1, pp. 96-112

ISSN 3008-2021 doi.org/10.61071/JDP.2685

Considerations for Effective Online Assessment in the Age of Artificial Intelligence

Samuel Munalula Munjita

Department of Biomedical Sciences, School of Health Sciences, University of Zambia, Lusaka, Zambia

samuel.munjita@unza.zm – corresponding author

https://orcid.org/0000-0002-8638-9373

Phebby Mwangala Kasimba

Department of Pre-Medical Sciences, Lusaka Apex Medical University, Lusaka, Zambia

https://scholar.google.com/citations?hl=en&user=NDOYTEMAAAAJ

Abstract: Online assessment offers flexibility, scalability and rapid feedback, yet generative artificial intelligence (GenAI) has weakened the assumption that a submitted product reliably represents a student's independent learning. This critical essay argues that higher education should shift from product-centred assessment to a process-plus-performance model that makes reasoning, evidence selection, revision, feedback use, ethical judgement and intellectual ownership visible. Drawing on process-writing theory, constructivism, constructive alignment, assessment for learning, self-regulated learning, the Community of Inquiry, authentic assessment and Universal Design for Learning, it examines the limitations of take-home essays, remote multiple-choice tests, AI detection and surveillance. It proposes staged assignments, critical evaluation and transparent disclosure of AI use, authentic and localised tasks, reasoned multiple-choice questions, brief oral verification and targeted secure assessment at key progression points. Because process records can also be fabricated, trustworthy judgement should triangulate written development with live explanation or practical performance. This layered approach supports academic integrity without treating surveillance as the default, while strengthening critical thinking, AI literacy, inclusion and professional accountability.

Keywords: online assessment; artificial intelligence; process pedagogy; authentic assessment; oral examination; academic integrity; universal design for learning

Received: 02.07.2026. Accepted and published: 27.08.2026

Citation:
Munjita, S. M., & Kasimba, P. M. (2026). Considerations for Effective Online Assessment in the Age of Artificial Intelligence. *Journal of Digital Pedagogy, 5*(1) 96-112. Bucharest: Institute for Education. https://doi.org/10.61071/JDP.2685

1. Introduction

Assessment is not merely the collection of student work. It is an inferential process through which educators evaluate what a learner knows, understands and can do (Bennett, 2011). That evaluation is defensible only when the evidence is sufficiently valid, reliable, inclusive and attributable to the student. Generative artificial intelligence has destabilised this inference (Desai, 2025). A student may now generate a coherent essay, rewrite weak prose, synthesise readings, construct a case analysis, produce computer code or formulate an examination response within minutes. The educational problem is therefore deeper than plagiarism. It concerns whether the visible product still represents the intellectual activity and competence that the assessment was intended to measure (Lodge et al., 2023; Perkins, 2026; Pérez-Pérez et al., 2026).

This disruption should not be treated as evidence that online assessment has failed. Digital assessment can support large cohorts, shorten feedback cycles, enable frequent low-stakes practice, provide learning analytics and widen access for geographically dispersed learners (Stamer et al., 2026). Research in engineering education confirms advantages in efficiency, scalability and speed of feedback, although these benefits are accompanied by challenges involving technical reliability, academic integrity and the assessment of complex problem-solving processes (Akhtar & Perea-Borobio, 2026). The central question is therefore not whether universities should abandon online assessment, but how they can obtain trustworthy evidence of learning without sacrificing accessibility, educational value or students' preparation for AI-enabled workplaces.

The GenAI-assessment relationship has been described as a wicked problem because each attempted solution generates trade-offs and there is no final technical fix that satisfies all disciplines, students and institutional contexts (Corbin et al., 2025). Prohibiting AI may protect selected forms of independent performance, but it can also ignore legitimate workplace practices and drive use underground. Permitting AI without redesign may reward cognitive outsourcing and widen inequalities in access and digital literacy (Capraro et al., 2024). Surveillance and AI-detection systems may deter some misconduct, yet they are fallible, intrusive and potentially damaging to trust. Authentic assessment is valuable, but AI can also generate realistic business plans, clinical discussions and reflective narratives. Consequently, institutions need a pedagogical response rather than an arms race between increasingly capable AI systems and increasingly restrictive controls (AlBlooshi, 2026).

This essay advances a process-over-product argument. Students should be assessed not only on the final artefact, but also on how they framed the problem, located and evaluated evidence, used or rejected AI assistance, responded to feedback, revised their work and defended their judgement. Process pedagogy is both preventive and developmental (Anson, 2014): it reduces the value of submitting an outsourced product while making the assessment itself a sequence of learning opportunities. Nevertheless, process is not a complete solution because process records can be fabricated and because professional competence often requires live performance. The more defensible formulation is therefore *process plus performance*: staged evidence of learning should be triangulated with authentic application, brief oral verification, practical demonstration and targeted secure assessment at consequential progression points (Flower & Hayes, 1981; Nicol & Macfarlane-Dick, 2006; Perkins, 2026).

Using data from a search of published literature in two major electronic databases (PubMed, and Google Scholar) and general google search, this discussion develops the process-over-product argument in six steps. It first examines why conventional take-home essays, remote multiple-choice examinations and surveillance-based controls provide increasingly weak evidence. It then establishes the theoretical basis for process pedagogy, including constructivism, self-regulated learning, formative assessment and the Community of Inquiry. The essay subsequently considers AI-integrated written assignments, authentic and resilient tasks, oral examinations, reasoned multiple-choice questions, Universal Design for Learning and the needs of resource-constrained contexts. It concludes by proposing a layered programme-level model and the institutional conditions required for sustainable implementation.

2. Online assessment before and after generative AI

The emergency transition during the COVID-19 pandemic accelerated digital assessment, often before institutions had adequate infrastructure, policy or staff preparation (Matsieli & Mutula, 2024). Earlier research documented concerns about identity verification, collaboration, unauthorised resources and the difficulty of invigilating students remotely (Gamage et al., 2020). These weaknesses did not begin with GenAI. Take-home assessments were already vulnerable to contract cheating, peer assistance and plagiarism, while remote tests could be compromised through messaging applications, shared screenshots or second devices. In programming education, for example, Hellas et al. (2017) identified help-seeking, collaboration and systematic cheating in take-home examinations, demonstrating that ambiguous boundaries between legitimate assistance and misconduct predated contemporary chatbots.

GenAI has changed the scale, speed and invisibility of assistance. Unlike conventional copying, AI produces novel text for each prompt, making similarity-based detection less useful. It also blurs the boundary between support and substitution (Arora et al., 2025). Language correction, brainstorming and explanation may scaffold learning; generating the central argument or analysis may replace it. The same action can have different educational meanings depending on the learning outcome. Using AI to improve grammar may be reasonable in a molecular biology assignment focused on data interpretation, yet inappropriate in an assessment designed to test scientific writing. Assessment design must therefore specify what intellectual work belongs to the student rather than treating all AI use as a single category (Khlaif et al., 2024; Perkins et al., 2024; Pérez-Pérez et al., 2026).

The transformation is also part of a longer history of educational technologies. Digital platforms, adaptive systems, automated feedback and intelligent tutoring have progressively changed how students access support and how institutions organise learning. AI is not an isolated interruption but a more powerful stage in this trajectory (Tan et al., 2025). The appropriate response is consequently adaptation with educational purpose: universities should decide which competencies must remain independently demonstrable, which competencies should include responsible AI use, and which assessment conditions provide credible evidence of each. This distinction is especially important in health professions, engineering and other safety-critical programmes. A graduate may appropriately use AI to search options or draft documentation, but professional responsibility still requires the ability to recognise an unsafe recommendation, interpret evidence, explain uncertainty and act without assistance when systems fail. In such contexts, assessment must establish both AI-augmented competence and independent foundational competence. The issue is not choosing between human reasoning and AI use; it is preventing AI from concealing the absence of reasoning that future practice requires (Mate & Weidenhofer, 2022; UNESCO, 2023; Perkins, 2026).

3. Why product-centred online assessment is no longer sufficient

3.1 Take-home essays and the authorship problem

Academic writing remains educationally important. It develops argumentation, synthesis, disciplinary communication and the organisation of complex ideas. The weakness lies not in writing itself but in treating the final document as sufficient evidence of learning. A product-centred assignment shows the outcome but obscures the route by which the student selected sources, tested interpretations, revised claims or rejected alternatives. GenAI widens this evidential gap because a fluent product can be generated without the reading, struggle and reflection through which learning normally develops (Flower & Hayes, 1981; Lodge et al., 2023; Fan et al., 2025).

Making a task more 'authentic' does not automatically solve the problem. A personalised case, reflective account or professional report may still be generated if the student supplies enough contextual information. Kirsanov et al. (2026) found that students supported real-world and data-based tasks, and such tasks can reduce superficial misuse by requiring interpretation of local evidence. However, their study also suggests that unclear rules and fear of penalties encourage strategic non-disclosure. Authenticity must therefore be combined with explicit expectations and personal accountability. The student should be required to explain why particular evidence was trusted, how an interpretation was reached and what limitations remain.

Detection provides weak foundations for this judgement. Human markers may mistake genuine student work for AI-generated text or fail to recognise extensively AI-mediated work. Automated detectors have varying error rates and can generate false accusations, particularly for multilingual writers or formulaic academic prose. A system that depends on detection shifts staff effort from teaching to investigation and can create a low-trust relationship in which students conceal even legitimate uses. The scoping review by Pérez-Pérez et al. (2026) shows that transparency is widely advocated but poorly operationalised: requirements are often vague, disclosures are frequently unverified and evidence about workload and acceptability remains limited. The implication is that institutions need assessable transparency rather than an unsupported declaration or an unreliable detector.

3.2 Remote multiple-choice examinations

Multiple-choice questions (MCQs) offer broad curriculum coverage, objective scoring and efficiency in large classes. Well-designed items can assess application and interpretation through clinical scenarios, graphs, laboratory results or engineering data. Their vulnerability arises when conventional one-best-answer questions are completed remotely without credible supervision. Students can collaborate, consult an AI system or distribute questions among a group. Randomising item order and drawing from question banks may slow this behaviour, but it does not prove independent reasoning (Mate & Weidenhofer, 2022; Akhtar & Perea-Borobio, 2026).

Technical restrictions can also distort the construct being measured. Lockdown browsers control the examination device but not a second telephone, an off-camera helper or printed material. Preventing backtracking, imposing severe time limits and displaying countdown timers may increase anxiety and impair normal problem-solving strategies. In technical disciplines, students may also be unable to show working or receive partial credit when the platform records only a final answer. Akhtar and Perea-Borobio (2026) report that engineering students valued rapid feedback but were concerned that digital formats sometimes concealed their method, even though method and logic are central to disciplinary competence. A test can therefore become more secure in appearance while becoming less valid educationally.

3.3 Surveillance, proctoring and the return-to-examination argument

One response to GenAI is to return to invigilated handwritten examinations. This approach has a legitimate role where unaided recall, calculation or judgement is an essential threshold competence. It offers stronger identity and environmental control than an unsupervised take-home task. Yet using it as the universal solution would narrow assessment diversity, privilege performance under timed conditions and assess students in a manner that may be less representative of contemporary professional work. It would also reverse gains in accessibility and flexibility achieved through digital education (Gamage et al., 2020; Lodge et al., 2023; Perkins, 2026).

Remote proctoring attempts to reproduce the examination hall through cameras, screen monitoring, identity checks and behavioural analytics (Malhotra & Chhabra, 2026). Such tools may be proportionate for selected high-stakes decisions, but continuous use raises privacy, data-protection, cultural and equity concerns (Heinrich, 2025). Proctoring must consider learner contexts and environment. A camera cannot reliably reveal a second device or a person outside its field of view, while video demands bandwidth and exposes private home environments (Malhotra & Chhabra, 2026; Heinrich, 2025). Students with unstable internet that makes video difficult to use, shared accommodation, disability or anxiety may be penalised for conditions unrelated to the intended learning outcome. Security is therefore not synonymous with validity. A defensible assessment must distinguish barriers that protect the competence standard from barriers that merely make participation more difficult (Gamage et al., 2020; CAST, 2024; Lodge et al., 2023; Heinrich, 2025).

The alternative is not unrestricted trust. Excessive reliance on unsupervised products can leave an institution unable to establish who performed the intellectual work. The defensible position lies between universal surveillance and unverified take-home assessment: security should be targeted according to risk, while most assessments should be redesigned to generate richer evidence of how learning occurred (Lodge et al., 2023; Corbin et al., 2025).

4. The theoretical case for process-over-product pedagogy

4.1 Process writing, constructivism and constructive alignment

Process-over-product pedagogy begins with a different conception of writing and learning. Flower and Hayes (1981) described writing as a recursive cognitive process involving planning, translating ideas into text and reviewing. Writers do not simply deposit finished knowledge on a page; they move repeatedly between problem definition, evidence, drafting and revision. When assessment recognises only the final artefact, it ignores much of the cognitive work through which understanding is built (de Jong & De Haro, 2026). Requiring proposals, evidence maps, drafts and revision rationales makes these activities visible and gives them educational value.

This position is consistent with constructivist accounts of learning. Learners actively construct meaning by connecting new information to prior understanding, testing ideas and negotiating interpretations rather than merely receiving correct answers (Do et al., 2023). GenAI can support this activity when it acts as a provisional interlocutor, but it can undermine it when the learner accepts fluent output without intellectual engagement. The assessment task should therefore require the student to transform, challenge and apply information, not merely reproduce it (Garrison et al., 2000; Zimmerman, 2002; Fan et al., 2025). Over-reliance on GenAI to generate knowledge without critiquing the output supports attributes of the banking model of education which treats students as empty accounts into which teachers deposit knowledge (Takona, 2025). Moreover, GenAI may overlook local contexts and learners' lived experiences. Using GenAI in this manner therefore conflicts with Freirean principles of dialogue, critical consciousness, active participation and the co-construction of knowledge, all of which position learners as reflective agents capable of questioning information, relating it to their context and using it to transform their social reality (Takona, 2025).

Constructive alignment strengthens the argument. Biggs (1996) proposed that learning outcomes, teaching activities and assessment should require the same kind of cognitive performance. If an outcome states that students will critically evaluate evidence, an assignment that can be completed by accepting an AI-generated synthesis is misaligned, even when the final prose appears excellent. The task should instead ask students to expose their criteria for judging evidence, compare competing explanations, revise an initial position and defend the final decision. Process evidence is not an additional administrative burden attached to the 'real' assignment; it is part of the aligned performance.

4.2 Assessment for learning, feedback literacy and self-regulation

Assessment for learning treats assessment as an intervention that shapes learning rather than a terminal act of measurement. Black and Wiliam (1998) showed the importance of evidence that helps teachers and students identify the gap between current and desired performance. Sadler (1989) similarly argued that improvement requires learners to understand the standard, compare their work with it and take action to close the gap. A single high-weight final submission provides little opportunity for this cycle. A staged assignment, by contrast, creates points at which misconceptions can be identified, feedback can be interpreted and revisions can be justified.

Feedback is effective only when students develop the capacity to use it. Carless and Boud (2018) describe feedback literacy as the understandings and dispositions needed to appreciate feedback, make judgements, manage affect and take action. Process pedagogy can assess these capacities directly. A revision commentary can require the student to identify which feedback was adopted, which was rejected and why. Peer review can develop evaluative judgement because students learn to recognise quality in the work of others and apply similar criteria to their own work (Boud & Molloy, 2013). The final grade then rewards improvement and decision-making, not only surface polish.

Self-regulated learning provides a further basis. Effective learners plan, monitor and evaluate their activity, adjusting strategies when understanding is weak (Nicol & Macfarlane-Dick, 2006; Zimmerman, 2002). GenAI can either support or short-circuit this cycle. Fan et al. (2025) found that access to GenAI improved essay performance but did not necessarily produce equivalent gains in knowledge or transfer, and it altered learners' metacognitive and self-regulatory processes. Shaw and Nave (2026), in a working paper, describe 'cognitive surrender' as the tendency to adopt AI output with insufficient scrutiny. These findings should be interpreted cautiously, particularly where evidence is emerging, but they reinforce an important design principle: assessment should require students to monitor the quality of AI assistance rather than merely benefit from it.

4.3 Community of Inquiry, dialogue and co-creation

Online learning is strengthened when assessment is embedded in a community rather than treated as an isolated transaction. The Community of Inquiry framework locates meaningful learning at the intersection of cognitive presence, social presence and teaching presence (Garrison et al., 2000). Cognitive presence develops through inquiry, integration and resolution; social presence enables learners to participate as real people; and teaching presence structures and facilitates the experience. Process-rich assessment can activate all three. Students can discuss preliminary interpretations, critique peers' evidence, receive targeted guidance and then revise their individual conclusions.

Student co-creation extends this logic. Inviting students to formulate questions, propose assessment criteria or analyse examples can deepen their understanding of disciplinary standards (Bovill, 2020). Creating a credible MCQ with plausible distractors, for example, requires more than recalling the correct answer. The student must anticipate misconceptions, distinguish closely related concepts and justify why alternatives fail. These activities work particularly well in a flipped design: students prepare before class, test questions with peers and use contact time for explanation and challenge. Co-creation should not transfer all responsibility to students, but it can make assessment criteria visible and position learners as participants in knowledge-building rather than passive recipients of grades.

Process pedagogy therefore serves three functions. It generates evidence of authorship and intellectual development; it creates repeated opportunities for feedback and self-regulation; and it makes disciplinary judgement an explicit object of learning. These functions explain why process is a more educationally defensible response to GenAI than simply making questions harder or increasing surveillance (Flower & Hayes, 1981; Nicol & Macfarlane-Dick, 2006; Perkins, 2026).

4.4 The theoretical case for process-over-product pedagogy: Complementarities, Tensions and Learner Agency

The eight theoretical perspectives discussed above are best understood not as separate or additive explanations, but as complementary frameworks operating at different levels of process-over-product pedagogy. Process writing and constructivism explain learning as recursive knowledge construction, while constructive alignment translates this conception into assessment design by ensuring that learning outcomes, activities and assessment demand the same forms of intellectual engagement (Biggs, 1996). Assessment for learning, feedback literacy and self-regulated learning are particularly closely related: formative assessment provides information about the gap between current and desired performance, but its value depends on students developing the capacity to interpret feedback, judge quality and act on it independently (Sadler, 1989; Nicol & Macfarlane-Dick, 2006; Carless & Boud, 2018; Greene, 2020). A potential tension arises when formative assessment becomes overly teacher-directed, since excessive reliance on external feedback may inhibit learner autonomy and evaluative judgement; consequently, feedback should progressively support self-assessment, peer judgement and independent decision-making (Boud & Molloy, 2013). This concern also resonates with Freire's *Pedagogy of the Oppressed*, particularly his critique of the banking model of education, in which learners are treated as passive or empty receptacles into which knowledge is deposited by an authority rather than as active participants in its construction (Freire, 1970; Takona, 2025). Applied to GenAI, an uncritical reliance on AI-generated explanations risks reproducing a similar relationship: the learner merely receives, stores and reproduces apparently authoritative knowledge without interrogating its assumptions, contextual relevance or accuracy. Such practice can diminish dialogue, critical consciousness and learner agency, especially when AI-generated knowledge is privileged over students' lived experiences and local contexts. A Freirean alternative is therefore a problem-posing and dialogic pedagogy in which students question information, relate it to their own social and disciplinary contexts, challenge competing interpretations and participate actively in constructing meaning rather than simply consuming ready-made answers (Freire, 1970).

Similarly, although self-regulated learning often emphasises individual planning, monitoring and reflection, the Community of Inquiry and co-creation perspectives highlight the socially mediated nature of learning through dialogue, peer interaction and shared regulation, suggesting that autonomy develops through, rather than in isolation from, collaborative engagement (Garrison & Akyol, 2015; Na et al., 2024). Constructive alignment and co-creation also create a productive tension: alignment requires clear and consistent learning outcomes and standards, whereas co-creation promotes student agency and participation in shaping learning activities and assessment processes (Bovill, 2020). These positions can be reconciled by maintaining essential disciplinary standards while allowing students meaningful influence over questions, examples, evidence, peer critique and interpretations of criteria. In this integrated perspective, process evidence should therefore not become merely a mechanism for surveillance or authorship verification; rather, it should make students' reasoning, questioning, dialogue, revision and intellectual decision-making visible. Process-over-product pedagogy is thus theoretically defensible because it connects knowledge construction, formative feedback, evaluative judgement, self- and co-regulation, dialogue, critical consciousness and learner agency, enabling students to demonstrate not only what they produce but also how they question knowledge, develop their position and take intellectual responsibility for their scholarly work.

5. Designing process-rich assignments that integrate AI critically

A process-rich written assignment should be designed as a coherent sequence rather than as a final essay preceded by ungraded paperwork. A practical structure may include: (1) a problem statement or proposal; (2) an annotated evidence table or data plan; (3) a provisional thesis or solution; (4) peer or lecturer feedback; (5) a draft or worked example; (6) a revised final product; and (7) a reflective account explaining key decisions. Marks should be distributed across the sequence so that the process is consequential. The precise stages should match the discipline: a laboratory report might include protocol justification, raw data and analysis decisions, whereas a clinical case might include differential reasoning and response to new information (Flower & Hayes, 1981; Carless & Boud, 2018; Lodge et al., 2023).

Where AI is permitted, it should become a visible object of critical evaluation rather than a hidden substitute for learning. Students can be asked to generate an initial response, compare it with course evidence, identify omissions or hallucinations, test references, revise the answer and explain why the final version is more defensible. In this design, the quality of the prompt matters, but the assessed competence is not prompt cleverness alone. It is the student's ability to recognise what the output assumes, where it is weak and how evidence changes the conclusion. This approach aligns with educators' reported use of GenAI-assisted assignments in which students brainstorm with AI but must critique, develop and document the generated material (Khlaif et al., 2024).

Different tasks require different levels of AI involvement. The AI Assessment Scale provides a useful language for distinguishing tasks in which AI is prohibited, used only for planning, used for collaboration, integrated into the production process or explored as part of the learning outcome (Perkins et al., 2024). A programme need not adopt the scale mechanically, but it should communicate boundaries at task level. Phrases such as 'AI is allowed' or 'AI is prohibited' are too broad unless students know whether they may brainstorm, translate, edit language, analyse data, generate code or compose substantive arguments.

Transparency must also be assessable and proportionate. A declaration should state the tool, purpose and parts of the work affected. For more substantial use, students may submit selected prompts and outputs, annotate changes or include a short verification log. However, requiring exhaustive interaction histories can create privacy concerns, excessive workload and incentives to conceal use. Pérez-Pérez et al. (2026) conclude that transparency is more feasible when requirements are explicit, aligned with pedagogy and non-punitive. The rubric should therefore reward critical disclosure and verification rather than treating every declared use as suspicious.

Process evidence remains vulnerable. A student could use AI to generate an artificial sequence of drafts or write a reflective commentary after the fact. This limitation should be acknowledged rather than obscured. Random formative checkpoints, short in-class tasks, version histories, targeted questions about specific decisions and brief oral verification can test whether the submitted process corresponds to actual understanding. The purpose is triangulation, not forensic reconstruction of every keystroke. Process data should be collected only when it contributes to learning or assurance; otherwise, it risks becoming another form of surveillance (Pérez-Pérez et al., 2026; Lodge et al., 2023; Perkins, 2026).

6. Authentic and resilient assessment

Authentic assessment is often proposed as the answer to AI because it asks students to perform meaningful tasks resembling professional practice. Wiggins (1990) argued that assessment should require learners to use knowledge effectively rather than merely recall it. A public-health student might interpret surveillance data and recommend an intervention; an engineering student might diagnose a system failure; a biomedical student might evaluate conflicting laboratory findings. Such tasks can promote transfer, judgement and professional identity (Ifelebuegu, 2023).

Authenticity, however, is not the same as resilience. GenAI can produce plausible reports, recommendations and role-based responses, especially when a student supplies detailed context. Perkins (2026) therefore extends authentic assessment through the concept of resilient assessment. A resilient task remains educationally meaningful while tying success to the student's process or performance and including a means of verifying genuine understanding. A project may permit AI during research and drafting, for example, but require the student to explain methodological choices, respond to an altered scenario or demonstrate the procedure live.

Data-rich and localised tasks can increase resilience. Kirsanov et al. (2026) report that students perceived real-world, data-based assessments as less conducive to simple AI shortcuts. The benefit does not arise because AI is unable to process data; rather, the student must connect the output to specific evidence, constraints and consequences. Individualised datasets, field observations, laboratory records or locally relevant cases can reduce the usefulness of generic answers. Nevertheless, educators should avoid artificial personalisation that adds complexity without advancing the learning outcome. The goal is meaningful specificity, not an endless attempt to write prompts that AI cannot answer.

Resilient assessment also recognises that graduates need two forms of capability. They must know how to work productively with AI, including prompting, verification, bias recognition and ethical use. They must also retain foundational knowledge and judgement when AI is unavailable, inappropriate or wrong. An assessment programme should therefore include tasks completed *with AI* and selected tasks completed *without AI*. This dual approach is more credible than either pretending AI will disappear from the workplace or allowing it to mediate every demonstration of competence (UNESCO, 2023; Perkins et al., 2024; Mate & Weidenhofer, 2022).

7. Oral examinations and live verification

7.1 Why oral assessment is valuable

Oral assessment is particularly useful because it converts assessment from a static product into an interaction. The examiner can ask the student to clarify an argument, justify a source, interpret an unfamiliar result or explain how a changed assumption would affect the conclusion. Memorised or AI-generated material may support an opening response, but superficial understanding becomes more visible when follow-up questions require adaptation. Oral assessment therefore provides direct evidence of reasoning while developing communication skills required in professional settings (Perkins, 2026).

Hartmann (2025) demonstrates that oral examinations are most effective when they are not introduced as an emergency anti-cheating device at the end of a course. Using backward design, the course should be organised around the knowledge and verbal reasoning students will need to demonstrate. Regular discussion, social annotation, structured questioning and opportunities to think aloud prepare learners for the final interaction. Questions, criteria and permitted notes can be communicated in advance while follow-up prompts remain responsive. Hartmann's account also suggests that the total time required for short oral examinations can be comparable with, and in some cases lower than, the time spent grading large sets of essays.

Oral examinations need not replace writing. Writing supports extended argument, careful evidence use and disciplinary communication; oral dialogue tests whether the student owns and can mobilise that argument. A five- to ten-minute verification following an essay, project or recorded presentation may be sufficient. Questions should focus on high-value decisions: Why was one method preferred? Which evidence was least reliable? What alternative interpretation was rejected? How would the recommendation change if a key value differed? Which part of the AI output required the greatest correction? The oral component then functions as both assurance and further learning (Hartmann, 2025; Perkins, 2026).

7.2 Anxiety, fairness, reliability and scale

The case for oral assessment must address legitimate concerns. Students may experience anxiety, communicate in a second language, have speech or hearing impairments, or be unfamiliar with rapid verbal response. Individual examinations can demand staff time, and unstructured questioning may introduce assessor bias. These are not reasons to dismiss oral assessment, but they require design discipline (Perkins, 2026; CAST, 2024).

Fairness begins with alignment. If professional oral communication is a learning outcome, it should be assessed directly. If the intended outcome is conceptual understanding, unnecessary demands for polished public speaking should not determine the grade. Students should receive the format, rubric and sample questions in advance, participate in low-stakes practice and have reasonable time to process each prompt. Alternatives may include a pre-recorded explanation followed by a shorter live discussion, additional response time, assistive technology, a quiet one-to-one setting or other adjustments consistent with the competence standard (Biggs, 1996; CAST, 2024; Perkins, 2026).

Reliability can be strengthened through structured question sets, common scenarios, scoring anchors, assessor training and moderation. Some variation is desirable because responsive follow-up is the source of validity, but the core domains and difficulty should be comparable. Recording a sample of sessions, using two assessors at critical points or conducting calibration exercises can reduce inconsistency. Perkins (2026) proposes that scaling should be addressed through strategic use, parallel delivery, structured formats and programme-level selection rather than assuming that every task requires a lengthy individual viva.

Anxiety should also be interpreted carefully. Assessment should not create avoidable distress, but all discomfort is not evidence of unfairness. Explaining ideas under questioning is a genuine professional competence in many fields. Preparation, predictability and supportive routines can convert anxiety into manageable challenge. The appropriate response is to scaffold the performance and remove irrelevant barriers, not automatically to remove an essential outcome (Hartmann, 2025; Perkins, 2026).

8. Rethinking multiple-choice and digital examinations

The process-over-product argument does not require abandoning MCQs. Their efficiency and diagnostic value remain important, especially in large cohorts. The priority is to redesign them so that they reveal reasoning. Two-tier questions ask students first to select an answer and then to select or construct the justification. Other formats require students to identify why distractors are wrong, rank alternatives, select all defensible options, state confidence or specify what additional evidence would resolve uncertainty. Such items reduce successful guessing and reveal misconceptions that ordinary one-best-answer questions conceal (Anderson & Krathwohl, 2001; Bolton et al., 2024; Mate & Weidenhofer, 2022).

In biomedical sciences, for example, an item can present symptoms and laboratory results, ask for the most likely diagnosis, and then require the student to identify the decisive finding, explain why a plausible alternative is inconsistent and select the next test. In engineering, an item can ask for a fault diagnosis and then require the sequence of reasoning or the assumption on which the answer depends. Digital platforms can automate the first tier while using short constructed responses or structured reasoning options for the second (Bolton et al., 2024; Mate & Weidenhofer, 2022).

Reasoned MCQs are educationally stronger but not AI-proof when unsupervised. A student can submit the entire item to a chatbot. Their greatest value is therefore formative and diagnostic: they can support retrieval practice, immediate feedback and misconception analysis. For high-stakes assurance, they should be completed in a controlled environment, combined with a practical task or sampled through oral verification. Randomisation is defensible only when question pools are equated for content, difficulty and cognitive demand (Bolton et al., 2024; Lodge et al., 2023).

Digital assessment should also recognise process in numerical and technical subjects. Stepwise answer fields, scanned workings, annotated calculations and partial-credit rules allow the mark to reflect method rather than only the final value. Akhtar and Perea-Borobio (2026) show that students want digital assessment to function as a meaningful checkpoint in learning, not merely as an efficient grading mechanism. Rapid feedback is useful, but it should explain misconceptions and direct revision rather than simply report correctness.

9. Universal Design for Learning, fairness and resource-constrained contexts

Assessment integrity and inclusion are sometimes presented as competing priorities, but exclusion can itself undermine validity. A mark should reflect the intended competence rather than the student's internet stability, device quality, home environment or familiarity with the learning-management system. This issue is particularly important in resource-constrained contexts, where students may depend on mobile telephones, intermittent data, unstable electricity or shared facilities. Evidence from blended-learning implementation in Zambia indicates that connection failure, power interruptions, mobile limitations and inadequate digital induction can prevent otherwise capable students from completing timed tasks (Limbumbu et al., 2022; Mabotha & Ngcamu., 2026; Munjita et al., 2026; Murillo-Jiménez et al., 2025).

Universal Design for Learning (UDL) offers a principled response. The UDL Guidelines emphasise multiple means of engagement, representation, and action and expression (CAST, 2024). In assessment, this means clear and accessible instructions, compatible documents, captions and descriptions for multimedia, keyboard navigation, opportunities to practise unfamiliar formats and reasonable options for demonstrating learning where the medium is not itself the outcome. UDL does not require lowering standards. It requires separating the competence standard from incidental barriers.

The same reasoning applies to oral assessment. When oral communication is not central, a recorded response plus short live verification may be appropriate. When safe laboratory performance is the outcome, practical demonstration remains necessary, but the instructions, environment and assistive arrangements can be made accessible. Similarly, a tightly timed remote test should be used only when speed is relevant. Automatic saving, a tested mobile interface, a reasonable completion window, a clear technical-failure procedure, alternative contact routes and an ungraded practice assessment should be standard features (CAST, 2024; Akhtar & Perea-Borobio, 2026; Perkins, 2026).

AI itself introduces equity questions. Students differ in access to paid models, high-speed connectivity, specialised software, English-language fluency and prior prompting experience. An assessment that assumes unrestricted AI use may unintentionally reward financial advantage or digital confidence. Institutions should provide access to approved tools where AI is required, offer AI-literacy instruction and design tasks so that premium features do not determine grades. Transparent rules must also recognise that assistive technologies and language-support tools may be essential accommodations rather than unfair advantages (UNESCO, 2023; Perkins et al., 2024; Khlaif et al., 2024).

Video monitoring should therefore be used sparingly and proportionately. It may be justified at selected certification points, provided students understand how data will be used and have access to an alternative venue. It should not become the default response to every quiz. In low-resource settings, a modest supervised assessment centre or approved regional partner may provide stronger assurance and greater fairness than attempting continuous remote surveillance of every learner (Gamage et al., 2020; CAST, 2024; Lodge et al., 2023).

10. A layered programme-level model of trustworthy assessment

No single assessment method can simultaneously maximise authenticity, accessibility, efficiency, security and depth. Trustworthy evaluation should therefore be built across a programme through multiple complementary forms of evidence. Drawing on the principles identified in the literature, the authors developed the layered programme-level model presented in Table 1 to illustrate how assessment can move from low-stakes learning activities to high-assurance progression decisions (Corbin et al., 2025; Lodge et al., 2023; Perkins, 2026).

Table 1

Illustrative layered model for online assessment in the age of AI

Assessment layer	Position on AI	Observable evidence of learning	Primary purpose
1. Formative and diagnostic	Allowed or encouraged with guidance	Quiz performance and correction of errors; explanations of answers; identification and correction of weaknesses in AI-generated responses; improvement between initial and repeated tasks	Practice, feedback and identification of misconceptions
2. Process-rich coursework	Bounded use specified by the task	Evidence of how the student developed the work: research question or proposal, source selection and justification, evidence map, selected AI interactions where relevant, draft-to-final changes, and explanation of how feedback was used	Reasoning, self-regulation, academic writing and transparent use of AI
3. Authentic performance	Used where it reflects professional practice	Application of knowledge to a new or local problem; interpretation of data; justified case or clinical decision; practical demonstration; completed project or professional product accompanied by a rationale for key decisions	Application, transfer and professional judgement
4. Oral or live verification	No live AI unless AI use is itself being assessed	Student explains or defends submitted work, answers follow-up questions, interprets unfamiliar information, responds to a changed scenario, or demonstrates a procedure without external assistance	Verification of understanding, authorship, adaptive reasoning and accountability
5. Secure progression point	Restricted or prohibited	Independently demonstrated knowledge or compe-tence under controlled conditions, such as correct responses in a supervised examination, successful completion of a practical or OSCE-style station, or defensible responses in a controlled oral assessment	Assurance of threshold competence for progression or certification

Note. The layers are complementary rather than sequential requirements for every assessment. Programmes should select and combine them according to the learning outcomes being assessed, the consequences of the assessment, the level of assurance required and available resources (Lodge et al., 2023; Perkins et al., 2024; Perkins, 2026).

The strength of the model lies in triangulation. An essay demonstrates extended argument; a process portfolio shows development and feedback use; an oral question tests ownership and adaptability; a practical activity demonstrates performance; and a controlled assessment verifies selected unaided competence. Weakness in one source of evidence can be interpreted alongside the others. The institution is no longer required to make an implausible claim that one unsupervised product proves everything (Lodge et al., 2023; Perkins, 2026). Programme mapping is essential. Educators should identify where each major learning outcome is introduced, practised and assured. Not every module needs an oral examination or secure venue, but every consequential competence should eventually be demonstrated under conditions appropriate to its risk. In health programmes, secure points may include progression into clinical or laboratory practice, capstone decisions and final competency assessments. Low-stakes activities can remain flexible and AI-supported because their primary purpose is learning (Biggs, 1996; Lodge et al., 2023; Mate & Weidenhofer, 2022).

This approach also manages workload more rationally. Staff time is concentrated where professional judgement adds the most value: feedback on pivotal process stages, oral sampling of significant work and assurance at threshold points. Automated marking can still support diagnostic quizzes and routine knowledge checks. The goal is not to replace efficient digital assessment with labour-intensive methods, but to prevent efficiency from becoming the dominant criterion for educational quality (Akhtar & Perea-Borobio, 2026; Hartmann, 2025; Perkins, 2026).

11. Institutional conditions for implementation

Assessment reform cannot be left to individual lecturers acting in isolation. Students encounter assessment across programmes, and inconsistent rules encourage confusion and concealment. Institutions need clear, discipline-sensitive policies specifying acceptable AI use, disclosure expectations, data protection, consequences of misconduct and routes for appeal. Policy should distinguish unauthorised substitution from legitimate support and should be translated into task-level instructions rather than remaining only in an academic-integrity handbook (Pérez-Pérez et al., 2026; Khlaif et al., 2024; Lodge et al., 2023).

Staff development is equally important. Educators require AI literacy, assessment-design expertise, confidence in oral and performance assessment, and strategies for inclusive implementation. Khlaif et al. (2024), using the Unified Theory of Acceptance and Use of Technology, found that adoption is influenced by perceived usefulness, ease of use, social influence and enabling conditions. Asking lecturers to redesign assessment without workload recognition, technical support, exemplars or peer communities is unlikely to produce sustainable change. Institutions should support pilots, moderation groups, shared question banks and communities of practice.

Students need induction in both digital assessment and AI ethics. They should practise platform functions before marks are attached, learn how to cite or declare AI use, understand hallucination and bias, and recognise when professional responsibility requires independent judgement. Clear expectations can reduce the penalty anxiety identified by Kirsanov et al. (2026) and make disclosure a normal component of scholarly practice rather than an admission of wrongdoing.

Technical and governance arrangements must support the pedagogy. These include accessible low-bandwidth platforms, automatic saving, contingency procedures, device-loan or supervised-access options, privacy governance for recordings and proctoring data, and secure storage with defined retention periods. Quality assurance should review whether AI conditions, rubrics and verification methods are aligned with learning outcomes and applied consistently. Workload models should recognise the time required for oral assessment, feedback and moderation (Akhtar & Perea-Borobio, 2026; Gamage et al., 2020; CAST, 2024). Finally, reform should be iterative. Because AI capabilities and student practices change rapidly, institutions should pilot designs, collect student and staff feedback, analyse grade patterns and revise. Corbin et al. (2025) argue that wicked problems require informed experimentation rather than a claim to final resolution. A learning institution should therefore treat assessment reform as an ongoing programme of inquiry, with explicit evaluation of educational benefit, integrity, equity and feasibility.

12. Discussion

The evidence and theory considered in this essay appear to support neither uncritical technological optimism nor technological panic. GenAI may widen access to explanation, stimulate ideas, provide rapid feedback and help prepare students for emerging professional practice (UNESCO, 2023; Khlaif et al., 2024). However, it may also enable students to submit work without fully engaging in the reading, uncertainty, revision and judgement that the assignment was intended to cultivate (Fan et al., 2025). We therefore argue that the central educational question is not simply whether an AI tool was used, but whether the student remained cognitively and ethically responsible for the work produced. Four broad responses can be compared. (1) Prohibition and a return to invigilated examinations may provide relatively strong assurance for selected foundational competencies (Lodge et al., 2023), but we also contend that universal reliance on such approaches could narrow the curriculum and privilege memory and speed over more authentic forms of professional performance. (2) Surveillance and detection may deter some forms of misconduct, but they cannot reliably establish authorship and may introduce privacy, equity and trust concerns (Corbin et al., 2025; Perkins, 2026). (3) Unrestricted AI integration may better reflect emerging workplace practices and improve the quality of student products, but without assessment redesign it could also amplify inequalities in access and encourage cognitive outsourcing (Khlaif et al., 2024; Fan et al., 2025). (4) On balance, pedagogical redesign based on process evidence, authentic performance, transparency and proportionate verification offers the most defensible response because it seeks to address learning and integrity together rather than as separate problems (Lodge et al., 2023; Perkins, 2026).

Within this position, process-over-product pedagogy is proposed as a particularly promising organising principle because it can make important aspects of learning activity more visible and assessable. This argument is supported by process-writing theory, which treats writing as recursive (Flower & Hayes, 1981), and by constructivist approaches that emphasise active meaning-making. It is also consistent with assessment for learning, which uses evidence to support improvement (Black & Wiliam, 1998), self-regulated learning, which emphasises planning, monitoring and reflection (Zimmerman, 2002), and feedback literacy, which develops students' capacity to judge and revise their work. We suggest that process-oriented assessment may help address the risk of metacognitive laziness by requiring students to examine, evaluate and justify their use of AI rather than simply accept its outputs (Fan et al., 2025).

At the same time, we do not regard process evidence as inherently authentic or trustworthy. AI can generate outlines, drafts, feedback responses and reflective accounts, while version histories may be manipulated and extensive logging may become intrusive (Pérez-Pérez et al., 2026). Process evidence should therefore be treated as one source of evidence rather than as a definitive audit trail. Lodge et al. (2023) similarly emphasise the need to rethink assessment assurance in ways that do not depend on a single indicator of student learning or authorship. Evidential value of process documentation is strengthened when it is selectively sampled and triangulated with short oral questions, in-class checkpoints, practical application or other forms of direct performance. From this perspective, the model advanced in this essay is more of a process plus performance rather than process alone (Perkins, 2026).

A similar qualification applies to authentic assessment. Real-world tasks may increase motivation, relevance and opportunities for application, but authenticity alone should not be assumed to protect against inappropriate dependence on AI. Kirsanov et al. (2026) highlight the growing difficulty of inferring genuine understanding from the quality of an AI-assisted product alone. It can be argued that authentic assessment becomes more trustworthy when it incorporates accountability by requiring students, where appropriate, to explain, adapt, justify or perform what they have submitted. Perkins' (2026) concept of resilient assessment is useful in this regard because it preserves the educational value of authentic tasks while recognising the possibility that a convincing product may be produced without an equivalent level of understanding.

Oral assessment may provide a particularly direct form of accountability, but we caution against treating it as an uncomplicated solution. Unstructured viva examinations may reproduce bias, disadvantage anxious or multilingual students and place substantial demands on staff time. Hartmann (2025) indicates that these limitations can be mitigated through careful design, structured questioning, clear criteria and opportunities for students to become familiar with the assessment format. Perkins (2026) similarly argues for proportionate and strategically placed verification rather than universal high-security assessment. Therefore, targeted use, such as a brief oral defence attached to major coursework or deployed at critical progression points, rather than lengthy oral examinations after every assignment is favoured in this discussion.

The layered model proposed in this essay may be particularly useful for resource-constrained universities because it avoids framing assessment as a choice between trusting all remote submissions and investing in comprehensive surveillance systems. Lodge et al. (2023) argue for assessment systems that distribute assurance across different forms and stages of assessment rather than relying on a single high-security response. Under the model proposed here, flexible formative assessment and process-rich coursework can occur online, while limited secure facilities can be reserved for higher-consequence decisions. Low-bandwidth design, practice opportunities and alternative venues may also support inclusion, particularly in settings where connectivity and infrastructure remain uneven (Munjita et al., 2026). Such an approach should not necessarily be viewed as a weaker integrity model; rather, it can be understood as a more risk-sensitive allocation of assessment resources, particularly where institutional capacity is constrained (Akhtar & Perea-Borobio, 2026).

The model also offers a broader way of conceptualising academic integrity. UNESCO (2023) emphasises responsible, transparent and critically informed use of GenAI rather than reliance on prohibition alone. Building on this position, we suggest that integrity should encompass honest attribution, responsible use of assistance, verification of evidence, willingness to acknowledge uncertainty and ownership of professional decisions. Students who disclose how AI influenced their work, identify its limitations or errors and defend their final judgement may therefore demonstrate forms of integrity that cannot be captured solely through the absence of detected misconduct (Pérez-Pérez et al., 2026). At the same time, transparency is not sufficient in itself. Kirsanov et al. (2026) caution against equating disclosure of AI use with evidence of genuine intellectual engagement; a declaration that AI was used cannot by itself demonstrate that meaningful learning occurred.

A further concern is the possibility that the educational benefits and risks of GenAI will be distributed unevenly. Fan et al. (2025) suggest that students with stronger disciplinary and metacognitive resources may be better positioned to use GenAI productively, whereas less-prepared learners may be more vulnerable to accepting plausible but weak or inaccurate outputs. Differences in access to premium tools, connectivity and language resources may compound these disparities (UNESCO, 2023). Khlaif et al. (2024) similarly draw attention to contextual and infrastructural inequalities that shape students' capacity to benefit from emerging technologies. Therefore, we suggest that process-oriented pedagogy may mitigate some of these risks when it explicitly teaches prompt evaluation, verification and critical judgement, although institutional access and support remain necessary. In our view, assessment reform should be accompanied by AI literacy and equitable provision rather than limited to changes in assessment instructions alone.

Finally, we are of the belief that assessment design is closely intertwined with curriculum design because the distribution of marks influences where students direct their effort. Biggs (1996) demonstrated through constructive alignment that assessment strongly shapes the learning activities students perceive as necessary for success. Similarly, Black and Wiliam (1998) showed that assessment can be used to direct and improve learning rather than merely measure its final outcomes. Where grades reward primarily polished final products, students may be encouraged to optimise those products using whatever tools are available. In contrast, when assessment also rewards evidence selection, revision, critical AI use, explanation and application, it may encourage greater engagement with those learning processes. This position is also consistent with Nicol and Macfarlane-Dick's (2006) argument that assessment should develop students' capacity to monitor, judge and regulate their own learning. As a result, we contend that an educationally stronger response to GenAI is not simply to develop more sophisticated mechanisms for identifying misconduct after it occurs, but to design assessment systems in which the desired processes of learning, judgement and accountability are themselves integral to successful performance.

13. Conclusion

Generative AI should not push universities towards an endless contest of detection and surveillance, but towards a more purposeful understanding of assessment itself. The most credible assessments will be those that make learning visible, require students to exercise judgement, and hold them accountable for explaining and applying what they know. The goal is not to eliminate AI from education, nor to accept every AI-assisted product as evidence of competence, but to design assessment that develops independent thought, responsible technology use and genuine intellectual ownership. In the age of AI, academic rigour will depend less on controlling every tool students use and more on ensuring that they can think critically, act ethically and demonstrate understanding that is unmistakably their own.

Declarations

Funding. No external funding was received for this work.

Conflict of interest. The authors declare that they have no conflicts of interest.

Ethics approval and informed consent. Not applicable. This critical essay did not involve human participants, identifiable personal data or animal research.

Data availability. No new datasets were generated or analysed for this article.

Author contributions. S.M.M. conceptualised the article and prepared the original draft. S.M.M. and P.M.K. reviewed and edited the manuscript. Both authors approved the final version.

Use of generative AI. Generative AI was used for language refinement, structural editing and reference-list organisation. The authors reviewed and verified the final text and accept full responsibility for its accuracy, originality and integrity.

References

Akhtar, Z., & Perea-Borobio, E. (2026). Navigating the digital shift: Investigating the viability of digital assessments for electrical and electronic engineering students. *European Journal of Engineering Education, 51*(3), 714–733. https://doi.org/10.1080/03043797.2025.2523549

AlBlooshi, S. (2026). Artificial intelligence in higher education, opportunities, and challenges: A review. *Frontiers in Education, 10*, Article 1683968. https://doi.org/10.3389/feduc.2025.1683968

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Anson, C. M. (2014). Process pedagogy and its legacy. In G. Tate, A. R. Taggart, K. Schick, & H. B. Hessler (Eds.), *A guide to composition pedagogies* (2nd ed., pp. 212–226). Oxford University Press.

Arora, V., Thabane, A., Parpia, S., Calic, G., & Bhandari, M. (2025). Generative artificial intelligence models outperform students on divergent and convergent thinking assessments. *Scientific Reports, 15*, Article 36987. https://doi.org/10.1038/s41598-025-21398-4

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347–364. https://doi.org/10.1007/BF00138871

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bolton, P. F., Lehr, L. R., Simha, R., & Lawson, M. (2024). The justification effect on two-tier multiple-choice exams. In *2024 ASEE Annual Conference & Exposition proceedings*. American Society for Engineering Education.

Boud, D., & Molloy, E. K. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Bovill, C. (2020). Co-creation in learning and teaching: The case for a whole-class approach in higher education. *Higher Education, 79*(6), 1023–1037. https://doi.org/10.1007/s10734-019-00453-w

Capraro, V., Lentsch, A., Acemoglu, D., Akgun, S., Akhmedova, A., Bilancini, E., Bonnefon, J.-F., Brañas-Garza, P., Butera, L., Douglas, K. M., Everett, J. A. C., Gigerenzer, G., Greenhow, C., Hashimoto, D. A., Holt-Lunstad, J., Jetten, J., Johnson, S., Kunz, W. H., Longoni, C., . . . Van Bavel, J. J. (2024). The impact of generative artificial intelligence on socioeconomic inequalities and policy making. *PNAS Nexus, 3*(6), Article pgae191. https://doi.org/10.1093/pnasnexus/pgae191

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

CAST. (2024). *Universal Design for Learning guidelines version 3.0*. https://udlguidelines.cast.org. Accessed on July 19, 2026

Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, 1–17. https://doi.org/10.1080/02602938.2025.2553340 de Jong, E., & De Haro, S. (2026). Technological understanding: On the cognitive skill involved in the design and use of technological artefacts. *Synthese, 207*(3), Article 121. https://doi.org/10.1007/s11229-026-05503-2

Desai, H. (2025, May 19). *What's worth measuring? The future of assessment in the AI age*. UNESCO. https://www.unesco.org/en/articles/whats-worth-measuring-future-assessment-ai-age

Do, H. N., Do, B. N., & Nguyen, M. H. (2023). How do constructivism learning environments generate better motivation and learning strategies? The design science approach. *Heliyon, 9*(12), Article e22862. https://doi.org/10.1016/j.heliyon.2023.e22862

Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology, 56*(2), 489–530. https://doi.org/10.1111/bjet.13544

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*(4), 365–387. https://doi.org/10.2307/356600

Freire, P. (1970). *Pedagogy of the oppressed* (M. B. Ramos, Trans.). Herder and Herder. (Original work published 1968).

Gamage, K. A. A., de Silva, E. K., & Gunawardhana, N. (2020). Online delivery and assessment during COVID-19: Safeguarding academic integrity. *Education Sciences, 10*(11), Article 301. https://doi.org/10.3390/educsci10110301

Garrison, D. R., & Akyol, Z. (2015). Toward the development of a metacognition construct for the Community of Inquiry framework. *The Internet and Higher Education, 24*, 66–71. https://doi.org/10.1016/j.iheduc.2014.10.001

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Greene, J. A. (2020). Building upon synergies among self-regulated learning and formative assessment research and practice. *Assessment in Education: Principles, Policy & Practice, 27*(4), 463–476. https://doi.org/10.1080/0969594X.2020.1802225

Hartmann, C. (2025). Oral exams for a generative AI world: Managing concerns and logistics for undergraduate humanities instruction. *College Teaching*. Advance online publication. https://doi.org/10.1080/87567555.2025.2558563

Heinrich, E. (2025). A systematic-narrative review of online proctoring systems and a case for open standards. *Open Praxis, 17*(3), 485–499. https://doi.org/10.55982/openpraxis.17.3.836

Hellas, A., Leinonen, J., & Ihantola, P. (2017). Plagiarism in take-home exams: Help-seeking, collaboration, and systematic cheating. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 238–243). Association for Computing Machinery. https://doi.org/10.1145/3059009.3059065

Ifelebuegu, A. O. (2023). Rethinking online assessment strategies: Authenticity versus AI chatbot intervention. *Journal of Applied Learning & Teaching, 6*(2), 385–392. https://doi.org/10.37074/jalt.2023.6.2.2

Khlaif, Z. N., Ayyoub, A., Hamamra, B., Bensalem, E., Mitwally, M. A. A., Ayyoub, A., Hattab, M. K., & Shadid, F. (2024). University teachers' views on the adoption and integration of generative AI tools for student assessment in higher education. *Education Sciences, 14*(10), Article 1090. https://doi.org/10.3390/educsci14101090

Kirsanov, O., Kushwah, L., & Selvaretnam, G. (2026). Beyond detection: How students use—and hide—AI in online assessments and what authentic tasks can do about it. *Journal of Academic Ethics, 24*, Article 14. https://doi.org/10.1007/s10805-025-09691-3

Limbumbu, A. N., Kabwe, J. C., Kumwenda, A., Kasonkomona, P. C., Mwila, G., & Lubeya, M. K. (2022). Online learning during the COVID-19 pandemic: A qualitative study among final year medical students at the University of Zambia. *F1000Research, 11*, 1363. https://doi.org/10.12688/f1000research.124823.1

Lodge, J. M., Howard, S., Bearman, M., Dawson, P., & Associates. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality and Standards Agency. https://www.teqsa.gov.au/sites/default/files/2023-09/assessment-reform-age-artificial-intelligence-discussion-paper.pdf

Mabotha, P. A. P., & Ngcamu, B. S. (2026). Digital transformation in the higher education sector: A systematic literature review. *Administrative Sciences, 16*(1), Article 1. https://doi.org/10.3390/admsci16010001

Malhotra, M., & Chhabra, I. (2026). Ensuring academic integrity through automated online exam proctoring: A decade-long systematic review. *Discover Education, 5*, Article 207. https://doi.org/10.1007/s44217-026-01224-3

Mate, K., & Weidenhofer, J. (2022). Considerations and strategies for effective online assessment with a focus on the biomedical sciences. *FASEB BioAdvances, 4*(1), 9–21. https://doi.org/10.1096/fba.2021-00075

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), Article 819. https://doi.org/10.3390/educsci14080819

Munjita, S. M., Sitali, L., Samutela, M., Mushumba, M., Kasimba, P. M., Sichone, J., & Muleya, G. (2026). *Reflections on blended learning implementation in Zambia: An e-portfolio-informed autoethnographic and reflexive thematic analysis*. Unpublished manuscript.

Murillo-Jiménez, H., Centeno-Alarcón, M., Buele, J., & Yumbla, F. (2025). Analyzing barriers to the effective implementation of technological tools in inclusive education: A scoping review. *Frontiers in Education, 10*. https://doi.org/10.3389/feduc.2025.1687664

Na, C., Jeong, S., Clarke-Midura, J., & Shin, Y. (2024). Linking self-regulated learning to Community of Inquiry in online undergraduate courses: A person-centered approach. *Educational Technology Research and Development, 72*(6), 2895–2920. https://doi.org/10.1007/s11423-024-10380-y

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

Pérez-Pérez, I., González-Afonso, M. C., Plasencia-Carballo, Z., & Pérez-Jorge, D. (2026). Transparency mechanisms for generative AI use in higher education assessment: A systematic scoping review (2022–2026). *Computers, 15*(2), Article 111. https://doi.org/10.3390/computers15020111

Perkins, G. (2026). Resilient assessment in the age of AI: Authentic design and the case for verbal examinations in business education. *Assessment & Evaluation in Higher Education*. Advance online publication. https://doi.org/10.1080/02602938.2026.2644516

Perkins, M., Roe, J., & Furze, L. (2024). *The AI Assessment Scale revisited: A framework for educational assessment* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2412.09029

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144. https://doi.org/10.1007/BF00117714

Shaw, S. D., & Nave, G. (2026). *Thinking fast, slow, and artificial: How AI is reshaping human reasoning and the rise of cognitive surrender* [Working paper]. The Wharton School. https://doi.org/10.2139/ssrn.6097646

Stamer, T., Machinek, N., Steinhäuser, J., Goujard, M., & Flägel, K. (2026). The role of digital teaching methods in supporting practical skills training in the academic training of health professions: A scoping review. *BMC Medical Education, 26*, Article 361. https://doi.org/10.1186/s12909-026-08785-9

Takona, J. P. (2025). Transformative education: Paulo Freire's *Pedagogy of the oppressed* and its contemporary resonance. *Journal of Global Education and Research, 9*(1), 87–107. https://doi.org/10.5038/2577-509X.9.1.1351

Tan, X., Cheng, G., & Ling, M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence, 8*, Article 100355.

UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research. Accessed on July 19, 2026.

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation, 2*, Article 2. https://doi.org/10.7275/ffb1-mm19

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Authors' Biographies

Samuel Munalula Munjita, PhD in Infectious Diseases, is a lecturer at the University of Zambia with academic interests spanning biomedical sciences, infectious diseases, blended learning, online education, and assessment. He was the founding Team Lead of the BSc Biomedical Sciences Blended Learning Programme at the University of Zambia, where he contributed to the development and implementation of flexible approaches to teaching and learning for undergraduate students. Dr. Munjita is currently pursuing an MA in Online Teaching at The Open University, United Kingdom, further strengthening his expertise in digital and technology-mediated education. His current scholarly interests focus particularly on the challenges experienced by students studying through online and blended modes, including issues of access, engagement, learner support, and participation. He is also interested in the redesign of assessment for online learning environments, with particular emphasis on modified and authentic approaches that promote student engagement, academic integrity, and meaningful demonstration of learning.

Phebby Mwangala Kasimba, MSc in Physics Education, is a physics educator with experience spanning both secondary and higher education in Zambia. She began her professional career as a secondary school Physics teacher before transitioning to university teaching and currently teaches Physics to undergraduate students at Apex Medical University in Lusaka, Zambia. Her teaching includes the use of blended learning approaches that combine face-to-face instruction with online and technology-supported learning activities. Ms. Kasimba's academic interests centre on understanding the challenges affecting students participating in online and blended learning in Zambia, particularly within the context of higher education. She is also interested in identifying practical and context-appropriate solutions that can improve student access, participation, engagement, and learning outcomes in technology-mediated education.