Reimagining Success and Failure: Equitable Assessment Practices in an Age of Artificial Intelligence

Professor Mollie Dollinger[a] and Dr Juuso Henrik Nieminen[b]

[a] Curtin University, Australia; [b] Deakin University, Australia.

Abstract

With disruption comes the affordance to not only think about things differently, but to truly do things differently. In the context of Generative AI in higher education this holds unique promise on how students are assessed, particularly from the viewpoint of equity. For decades, assessment practices and designs have been flawed; compartmentalised, standardised, and disadvantageous to many students. In this conceptual paper, we explore how the emergence of GenAI may help to challenge entrenched and exclusionary notions of student success in two key dimensions: distribution of knowledge and agentic assessment. We argue that GenAI creates an opportunity to move away from assessment practices that sort, rank and grade students toward approaches that recognise diverse ways of knowing. Thus, rather than viewing GenAI as a threat to academic integrity, we propose it as a catalyst for more equitable assessment that values collaboration and student agency over individual performance and predetermined standards.

Practitioner Notes

1. Traditional assessment practices that sort and rank students systematically disadvantage equity students, despite well-intentioned accommodations and accessibility measures.
2. Generative AI challenges the individualistic assumptions underlying most assessments, creating opportunities to recognise knowledge as distributed across people, tools, and contexts.
3. Moving towards competency-based or pass/fail frameworks can shift focus from high-stakes moments to processual learning, potentially better supporting diverse student populations.
4. Agentic assessment enables students to participate in defining what 'success' and 'failure' mean, positioning them as partners rather than subjects of evaluation.
5. Rather than detecting cheating with GenAI, educators should focus on detecting learning by assessing how students navigate complex, resource-rich environments.

Keywords

Generative AI, student equity, assessment design, higher education, distributed learning, agentic assessment, grading

Introduction

The once sudden, and now steady, proliferation of GenAI systems across educational landscapes has rightfully caught many educators' attention. At the heart of these discussions has been the idea of change: change as both something that is happening to us but also change as something we are driving. These changes encompass a variety of measures, from incorporating GenAI into assessment through digital literacies (Crawford et al., 2023, Smith et al., 2025) to the alternative of designing for its exclusion through measures such as supervision, invigilation, or a return to traditional pen and paper (e.g., Foung et al., 2024; Rasul et al., 2024).

Yet most approaches, while practical, have largely focused on containment rather than reimagination. This is understandable in a time of rapid change, but there is an urgent need to use this disruption to critically challenge the very foundations of assessment. Decades of research highlighting the inequity of assessment remains largely unaddressed and, at times, even exacerbated. For example, in cases where students can now harness GenAI to complete their assessments, digital divides and inequity of tool access means that students from equity-deserving backgrounds are often disadvantaged. Alternatively, assessments which return to 'high stakes' formats, such as supervised or invigilated exams, or high-pressure oral examinations, once again reinforce the very inequities that many have long argued to address (Rae et al., 2025; Tai et al., 2023).

The challenge facing higher education is not that GenAI has broken assessment, it's that GenAI has made it impossible to ignore that assessment was already broken. Universities have long relied on assessment practices designed to sort and rank students, measuring their ability to reproduce knowledge under artificial constraints that bear little resemblance to how knowledge is actually created or used in professional contexts. GenAI simply exposes what was always true, that assessments privilege those who can perform in narrow, acontextual ways, and we've been assessing compliance with academic conventions as much as we've been assessing learning. The question is not how to restore the status quo, but whether that status quo is worth restoring, particularly for students whose ways of knowing have been systematically devalued by traditional assessment regimes. In this conceptual paper, we use this moment of disruption to ask: what does success look like now?

Assessment institutionally defines what counts as 'success' in higher education and, consequently, sets the boundaries for how we talk and think about 'failure'. And yet, the foundational assumptions driving these notions remain largely unquestioned. Two dominant approaches have emerged to address assessment inequities, each with significant limitations. The first focuses on identifying specific obstacles faced by particular student populations and proposing targeted accommodations to enhance access. While well-intentioned, this perspective remains fundamentally flawed as it not only focuses its attention on superficial factors, but it continues to place the burden of responsibility to adapt solely on students (Dollinger et al., 2024; Elkhoury et al., 2023). Students are singled out and identified as 'at risk', with accommodations designed to help them 'catch up' to predetermined standards. The second, and more recent, approach represents an important shift away from deficit-based thinking towards examining how dominant assessment practices systematically disable students (Nieminen, 2023). This helpfully redirects attention on structures, rather than individual inadequacies, but still fails to challenge the underlying problem of assessment: that for one student to succeed, someone else needs to fail.

Following McDermott and Varenne's (1995) explorations of success and failure in educational contexts, this raises questions about why higher education's cultural apparatus has become 'heavily invested in failure' (p. 11). Why is it necessary to produce and recognise winners and losers? In this study, we do not propose ideas to promote student success (and hinder failure) as typically understood. Rather, we aim to deconstruct the assessment structures that have fixated higher education on producing success and failure.

We seek to resist what has been called 'technological solutionism' (Taylor, 2021), or the urge to react to new technologies with a pragmatic, often simplistic, approach (Crawford et al., 2023). The disruption that GenAI has afforded the higher education sector presents a significant opportunity to reimagine assessment through an equity lens that has long awaited proper consideration. Unlike previous educational technologies that have been largely accommodated within existing structures (Bearman et al., 2023), GenAI's fundamental challenge to individual authorship and knowledge creation (Bearman & Ajjawi, 2023) makes it increasingly difficult to maintain current assessment practices without explicit justification.

> **Research Question:** How might Generative AI serve as a catalyst for reimagining assessment practices in higher education to challenge entrenched notions of success and failure, particularly for students from equity-deserving backgrounds?

To address this question, we begin by defining our key terms, including our framing of GenAI and equity. We then examine the pattern of surface changes and structural continuity that has characterised educational technology adoption in higher education, demonstrating why GenAI might represent a departure from this trajectory. The paper subsequently explores two key dimensions through which assessment practices might be reimagined: first, a shift from individualistic to distributed understandings of knowledge; and second, a move from predetermined standards toward agentic assessment that positions students as active participants in defining success. We then consider the implications of these shifts for grading practices, before concluding with reflections on the opportunities and risks of the transformations we propose.

Literature

The intersection of artificial intelligence and equity in higher education represents a rapidly evolving area of scholarship. Recent literature has documented how AI systems can both perpetuate existing educational inequities and create new forms of marginalisation, particularly through algorithmic bias and unequal access to digital resources (Ajjawi et al., 2023; Bong & Chen, 2024; Prinsloo & Slade, 2016). While some research examines how AI detection tools disproportionately impact non-native English speakers and students who are neurodivergent (Dawson, 2020; Pang et al., 2024), less attention has been paid to how GenAI might fundamentally reshape assessment paradigms in ways that could address longstanding inequities (Bearman & Ajjawi, 2023; Dawson et al., 2024). This conceptual paper contributes to this emerging conversation by exploring how GenAI's disruption of traditional assessment practices creates opportunities to reimagine success and failure through an equity lens.

Defining Our Key Terms

For the purposes of this paper, we use the term Generative Artificial Intelligence (GenAI) to refer to largely user-facing AI systems capable of producing content that was previously considered

uniquely human in origin. These systems can generate text, images, code, and other outputs through machine learning models, and are distinguished by their capacity to engage in seemingly creative and analytical tasks which challenge established notions of authorship and individual achievement that underpin many of our current assessment practices. Importantly, we approach GenAI not as a neutral tool but as a sociotechnical phenomenon that both reflects and shapes educational practices (Bearman & Ajjawi, 2023).

Key in our framing is also how we define equity. In this study, we focus on those students who have been systematically excluded or marginalised in higher education through both structural and cultural barriers (Naylor & Mifsud, 2020). This includes students from diverse backgrounds, first-in-family students, students with disabilities, those from lower socioeconomic contexts, and others whose ways of knowing and demonstrating knowledge have been unrecognised by dominant educational paradigms. Importantly, this framing seeks to move beyond attempts to categorise students, and we purposefully do not refer to 'equity students' as a blanket category. We further do not suggest that students are inherently 'lacking' (O'Shea et al., 2016). Rather, we position the importance of this topic in recognising that universities continue to forfeit significant intellectual and cultural resources when assessment practices fail to validate and harness the diverse knowledges these students bring to our learning communities. Put simply, universities lose out when we don't pursue real inclusion.

Surface Changes, Structural Continuity: Educational Technologies and Assessment

Educational technology has long promised to transform university assessment. Yet despite decades of innovation, the fundamental purpose of assessment remains largely unchanged: to sort, rank, and credential students (Bearman et al., 2023). New technologies tend to change how assessment looks without changing what it does. Online education illustrates this pattern. When universities first moved learning online, the promise was significant: expanded access for students who could not attend campus, and greater flexibility for those juggling work, caring responsibilities, or geographical distance (Guenther et al., 2023; Stone et al., 2019). For many students, these benefits have been real. Research shows that online learning can offer greater work-life balance and more control over when and how students engage with their studies (Gibson et al., 2022).

But implementing the same technologies that enable these possibilities has created new barriers. Students from lower socioeconomic backgrounds often lack reliable internet, appropriate devices, or quiet spaces to study at home (Ajjawi et al., 2023; Bong & Chen, 2024). And even when these students do access online learning, the platforms can work against them. Learning analytics systems track how students engage: when they log in, how long they spend on modules, whether they follow expected patterns. Students who log in at irregular hours, take longer to complete tasks, or pause their studies due to work or family demands are often flagged as "at risk" and directed toward remedial support, not because they are struggling academically, but because their circumstances do not match the assumptions built into the system (Prinsloo & Slade, 2016).

Why does this pattern persist? Technologies are typically designed to fit within existing assessment structures, not to challenge them. Assessment designers remain anchored by the assumption that their job is to sort students into categories of success and failure, even when their goal is to improve equity. This creates what Nieminen (2024) calls the paradox of inclusive assessment: the structural impossibility of achieving true inclusion within systems that require some students to fail for others to succeed (see also McDermott & Varenne, 1995).

Misconduct detection tools offer another example. These are now among the most widely used educational technologies in higher education. Rather than helping students learn about authorship and academic practice, they function primarily to identify and punish those who fall short, morally categorising struggling students as cheaters. As Dawson (2020) observed, the misconduct detection industry has grown so powerful that it now shapes university practices rather than serving them. Many educators spend considerable time on detection despite limited evidence that these tools actually reduce misconduct sector-wide (Stoesz, 2023). The dominance of these tools reveals just how deeply the sorting function of assessment is embedded in our institutions.

Given this history, it would be reasonable to assume that generative AI will simply repeat the pattern: promising transformation while reinforcing the status quo. Indeed, the rapid adoption of AI detection tools, despite their methodological limitations, false positives, and violations of procedural fairness (refer to Bassett et al., 2025), suggests we are already making familiar mistakes. If we approach GenAI through the same lens of surveillance, competition, and sorting, that is precisely what will happen. But this outcome is not inevitable. GenAI could represent a genuine departure from previous educational technologies, but only if we are willing to reconsider what assessment is actually for.

Reimagining Assessment

Reassembling success

What might emerge if assessment practices were reassembled anew, using GenAI as a catalyst for fundamental change? This section examines how reconceptualising success in the age of GenAI could become an inclusive endeavour. This analysis explores two key themes: distributed learning and agentic assessment. While these concepts have been discussed in educational literature prior to GenAI's emergence, they take on new significance and possibilities within the current digital context. Each theme offers a pathway for moving beyond traditional assessment paradigms toward more equitable practices that value diverse ways of knowing and learning.

From individualistic to distributed success/failure

Perhaps the most pervasive, and most harmful, myth underpinning many current assessment practices is that the most valid understanding of students' knowledge is that which resides neatly within an individual. This approach is rooted in a particular, individualistic view of the self that has pervaded human life in many contemporary contexts (e.g., Bellah et al., 1985). In assessment, considerable measures and practicalities are undertaken to ensure assessment represents the skills and competencies of individuals, not groups or communities. Consider the traditional testing hall, with students working in isolation on their examinations in ways that render 'knowledge' and 'knowing' into something seemingly individualistic. Yet in such situations, 'individuality' is in fact an accomplishment, something that has to be achieved and maintained by organising the examination hall physically and socially (e.g., by asking students to work alone and in silence, by asking students not to use tools beyond pens and papers, and by asking students to write *their* name in the test paper). This approach systematically privileges students whose strengths align with isolated, high-pressure performance while disadvantaging those who excel in collaborative, resource-rich, or extended-time environments. The effects of this individualistic paradigm are

clearly demonstrated in such closed-book, time-constrained examinations, which have been evidenced to show differential performance based on gender, socio-economic status, race and ethnicity, and disability (French et al., 2024). Yet when examination formats accommodate more distributed approaches to knowledge, these disparities diminish significantly (Spitzer et al., 2025).

The problem goes beyond measurement as well, because it also reflects a narrow, and artificial, understanding of knowledge and knowing. 'Knowing' is presented as an individualistic endeavour, as individual students consume and reproduce fragmented pieces of knowledge. Yet knowledge can also be seen as inherently social/distributed; constructed through complex interactions between people, tools, contexts, and methods (Fawns, 2022).

In our new digitalised context, the distributed nature of knowledge is far more difficult to ignore. When students for example use AI tools to brainstorm ideas, fact-check claims, or generate initial drafts, the boundary inevitably blurs between what was 'their' knowledge versus what was digitally crafted alongside them (Nieminen et al., 2025). This directly refutes the idea that we as educators can delineate what is 'inherently human', as the tools we use, created by humans, now too shape us as humans.

How could this impact assessment? In fact, some assessment practices already provide the blueprint, for example, open-book examinations, collaborative research projects, or simulations where students are using technology in an invigilated setting. These assessments prioritise the navigational capabilities students need, rather than static individual knowledge. However, perhaps the most significant opportunity lies in how we can reimagine group work, often a point of contention for many students, but especially equity students. Group work fails because such assessments attempt to impose individual accountability in what is collaborative (Rasooli et al., 2025). But what if group work instead aimed to help teams identify and leverage one another's unique strengths? How did the team synthesise their divergent perspectives? How did they work together, listen to one another, and co-create ideas? In this way, students could submit documentation of their collaborative processes, showcasing how ideas and artefacts evolved through interaction, evidencing their ability to function as an integrated knowledge network.

From predetermined to agentic success/failure

Another entrenched and flawed feature of university assessment practices is the relentless pursuit of standardisation. Assessment practices pursue standardisation rooted in industrial-era values where all students had to learn vocational skills in identical ways. Despite rhetoric about 'knowledge societies' and holistic development (i.e., Bildung philosophies, refer to Herdt, 2019), learning outcomes remain fixed and predetermined before students ever enter the classroom. Students have little say in what they're assessed on, how they're assessed, or when assessment occurs. This positions students as performers rather than drivers of their own learning journey, reinforcing power dynamics where educators hold absolute authority and students are measured against external standards that disregard their individual contexts.

Agentic assessment positions students as active agents empowered to make meaningful choices about what they learn, how they demonstrate knowledge, and when they engage with assessment tasks, yet without adhering to unhelpful individualism as discussed above. In other words, students are systematically and structurally enabled agency over what counts as 'success' and 'failure' in assessment. Assessment ecosystems must provide meaningful affordances for

students to exercise individual agency (Adie et al., 2018). Agency means recognising students' capacity for self-direction (with oneself and others) and validating their diverse ways of knowing within academic structures.

McArthur (2016) critiques traditional standardised assessment, noting it relies on procedural justice to ensure uniform processes deemed 'fair', yet actually privileges those already advantaged by the system. Standardised approaches force disadvantaged students to succeed only in predetermined ways. Assessment systematically marginalises students operating on nontraditional timescapes (Dollinger et al., 2025) or possessing cultural approaches to demonstrating knowledge that don't align with standardised processes. Students from non-dominant backgrounds must demonstrate knowledge in prescriptive ways that may conflict with their strengths or constraints. Students who are neurodivergent face penalties for non-conformist approaches; students with caregiving responsibilities encounter disadvantage through inflexible timelines. These practices persist despite research showing benefits when traditionally marginalised students can agentically represent their skills beyond traditional measures (Steele et al., 2024), and in ways determined by students themselves, too (Nieminen et al., 2025).

How might GenAI challenge this approach? GenAI excels at standardised, predetermined tasks, rendering traditional targets of measurement increasingly obsolete. This creates pressure to rethink what assessment actually values. Rather than asking what GenAI cannot do (a moving target), the more productive question is what we want students to learn through the assessment process itself. This reframes assessment around the navigational choices, reflections, and judgments students make, rather than the outputs they produce (Steele et al., 2024). The diversity emerges through the process itself, valuing the processual nature of learning over static outputs (Ryan & Judd, 2022). Digital technologies can support students' ongoing reflection, preserve their unique learning trajectories, and help them develop evaluative judgment in collaboration with others (Nieminen et al., 2025), although careful human oversight remains essential to ensure quality and avoid bias, particularly for students from non-dominant language backgrounds (Pang et al., 2024). Critically, GenAI might also provide an important mechanism for students to portray their competencies beyond tokenistic measures of marks. The value lies not in whether GenAI could produce such artefacts, but in students articulating their own development, making curatorial decisions, and demonstrating metacognitive awareness. Ajjawi and Boud (2024) highlight the potential for students to create 'persuasive portfolios' that combine university-validated achievements with student-curated portrayals of their developing professional personas.

Agentic assessment also allows for greater flexibility in assessment criteria and timelines, enabling students to potentially negotiate deadlines, choose how they demonstrate their learning, and participate in establishing what success looks like. This better reflects professional environments, where workers manage competing priorities, set their own timelines, and determine appropriate formats based on context and audience. Therefore, while some students undoubtedly benefit from more structured learning environments (e.g., McInnis, 2001), through agentic assessment universities may be able to more closely mirror workplace environments, where necessary skills include self-regulation and proactive contribution.

Implications for Grading

If we are to truly challenge entrenched notions of assessment practice, we must address the elephant in the room: grades. After all, the reward system of grades may be more powerful than any other element in higher education in denoting cultural values over what counts as 'success' and what is seen as 'failure'. While grades ostensibly arise as outputs of students' work, they function in many ways as the grounding principle driving our educational systems and, subsequently, educators' assessment design (as noted for a long time, such as by Hiner, 1973). And they remain deeply embedded in both the minds of students and educators, harking back to persistent assumptions that knowledge is not only objective, but quantifiable. This persists despite questions about the precision of numerical distinctions in grading practices and greater global attention to what is often called the 'ungrading movement' (Stommel, 2023). Emerging research suggests that removing traditional grading practices can be particularly beneficial to students who have been marginalised by educational systems, creating more equitable classrooms, rectifying power imbalances, and enabling greater student autonomy over their learning (Love & Daniel, 2025; Rapchak et al., 2023), although many studies have also emphasised ambiguous and paradoxical findings in a world driven by metrics and scores (Kjærgaard et al., 2023). Yet grades continue to serve as a stalwart in determining not only academic progression but broader notions of student success, despite evidence which suggests that students reflect with much more nuance and complexity than any mark can capture (Cachia et al., 2018; O'Shea & Delahunty, 2018). So how might GenAI serve as a catalyst for reimagining these deeply entrenched grading practices, particularly when we consider the distributed and agentic possibilities it presents?

To unpack this question, we have examined how GenAI disrupts two fundamental assumptions underlying traditional grading: individualism and predetermined standards. The implications for grading become particularly apparent when we consider the adoption of more distributed approaches to assessment. Currently, grades almost invariably focus on individualistic success and failure, even within group work assessments where students often receive individual marks, thus reiterating from the very onset that this remains fundamentally an 'individual' endeavour. Yet GenAI fundamentally challenges this individualistic approach by making visible the networked and collaborative nature of knowledge creation, where human creativity becomes entangled with AI capabilities in ways that resist simple attribution (Bearman & Ajjawi, 2023). Moving beyond this framing requires shifting focus away from the individual element toward the collective. Drawing on the idea of learning communities (e.g., Kilpatrick et al., 2003), this transformation involves guiding students through competency-based or pass/fail frameworks which emphasise the processual nature of learning (La Chimea et al., 2020), rather than relying on high-stakes moments in time. While such approaches are sometimes criticised for lacking detail, competency-based frameworks can be equally rigorous and specific as traditional marking rubrics (Ange et al., 2018; Spring et al., 2011). For instance, a competency-based approach might involve students demonstrating 'effective human-AI collaboration' through a portfolio that includes: their process documentation when working with GenAI tools, reflections on output verification strategies, and evidence of how they addressed AI bias in their work. Similarly, instead of grading a literature review with a numerical score, students might demonstrate competency in 'critical synthesis across sources' by showing evidence of how they evaluated conflicting viewpoints, integrated diverse perspectives, and developed their own analytical position. Rather than receiving a numerical grade, students would either demonstrate competency or continue developing until they

meet the standard. The key difference lies in how learning is framed through the lens of development and progression, where students harness assessment, and corresponding feedback, to reflect on their learning with respect to their 'own' achievement within broader ecosystems of learning communities (Baartman & Quinlan, 2024; Schuwirth & Van der Vleuten, 2011).

Alternatively, for institutions not ready to abandon numerical grading entirely, a hybrid approach focusing on navigational capabilities offers a middle path. Another possibility involves transforming marks through distributed learning while retaining numerical scoring but deriving scores not from individual performance but from students' navigation of complex and interconnected knowledge-rich environments. In this approach, the mark focuses not on the artefact produced but on students' navigational capabilities, considering how they interacted with digital and non-digital resources, how they harnessed the diverse and collective expertise of those around them, and how they drew upon their own histories, cultures, and contexts (Bearman et al., 2020). In practice, this might involve assessing a student's research project not solely on the final paper, but on their documented process: How did they leverage diverse sources (human experts, AI tools, databases)? How did they synthesise conflicting information? How did they adapt when initial approaches proved inadequate? The grade would reflect their navigational sophistication rather than just their written output. This becomes particularly relevant in the context of GenAI, where assessment could evaluate students' capacity to effectively orchestrate human-AI collaboration, assuring verifiable outputs and addressing inherent biases. As previously discussed by scholars both prior to and post AI, a focus on navigational and collaborative assessment would more closely align to most professional environments, where success is not an isolated performance but what one can accomplish within their context (Ajjawi & Boud, 2024; Jorre de St Jorre et al., 2021). Such a transformation would also recognise the multiplicities of knowledge, aligning to more non-Western collective understandings of knowing (Shen et al., 2025).

The relationship between grades and agentic assessment further reveals one of the most significant tensions in current practice. One of the biggest threats to agentic assessment lies in grades themselves, because as discussed, these are often tied to predetermined learning outcomes and expectations established long before students ever enter the classroom. This approach fundamentally divorces students from their own learning process. Rather than inviting them to reflect on what and why they should learn, it risks treating students as passive recipients of pre-packaged bits of knowledge where these decisions have already been made for them and remain non-negotiable. It becomes less surprising, then, when students disengage or when diverse students feel disconnected and disrespected within their learning environments.

Reframing this practice would position assessment as learning itself, promoting well-established mechanisms such as student self-grading that actively engage learners in discussing their own progress (Nulty, 2011; Sadler & Good, 2006). This could involve students submitting a learning reflection alongside their work, where they articulate what they learned, identify areas for growth, and propose their own grade with justification. Such approaches would position students as drivers of their educational choices rather than subjects of external judgment. In the context of GenAI, this becomes particularly vital for assessments that permit its use, as students must develop not only evaluative judgment in navigating human-AI collaboration but also the metacognitive awareness to understand their own learning processes within these networked

environments (Bearman et al., 2024). Agentic grading further aligns to the move towards more programmatic or course-wide assessment practices, as it could scaffold students' learning over time as their competencies evolve. In this way, GenAI tools could both catalyse change, as it stresses distributed learning, while supporting it, through helping students document their learning process and work towards 'assessment for distinctiveness' (Jorre de St Jorre et al., 2021).

It is worth noting that while much research remains needed to fully understand the impact of reduced grading in higher education, emerging evidence suggests complex dynamics at play. Students accustomed to grade-driven motivation may initially feel uncertain about their progress without numerical feedback, requiring scaffolded support to develop intrinsic learning motivation (Price et al., 2011). Future research should transparently and rigorously report the positive and negative (and anything in between and beyond) outcomes of trying to reform grading policies within an age of AI, particularly when equity is set as a goal. For example, some studies have shown that student motivation and even identity development suffer without traditional marks (Kjærgaard et al., 2023), although this perspective has been contested (Walden, 2022). The most significant obstacle likely stems from external pressures: future workplaces, professional bodies and further education institutions often rely on ranking systems to sort students (Spring et al., 2011). Additionally, nascent research indicates that students may require adjustment time when transitioning to ungraded or competency-based assessment, reflecting the deeply entrenched nature of traditional marking practices (Zarb et al., 2023). While such transformations require significant institutional commitment and faculty development, pilot programs could begin within individual courses to test these approaches before broader implementation.

What remains clear, regardless of these challenges, is that assessment structures must be fundamentally modified in the context of GenAI to recognise and respond to AI's transformative presence (Corbin et al., 2025; Ogunleye et al., 2024). Otherwise, we risk wasting resources on 'rigorous' grading practices – perhaps through anonymous grades, mark moderation and so forth – that may tell surprisingly little about our students within an age of AI, both for themselves and for our institutions. Worse still, without explicit attention to equity, these resources risk causing learning barriers for students and then blaming the victim ('here is your grade, D–'). Perhaps we should grade our grading systems rather than our students, particularly in these turbulent times?

Why does this reconceptualisation of assessment matter for learning and teaching practice? If universities are to genuinely serve their increasingly diverse student cohorts, the way we determine success and failure cannot remain tethered to assumptions that systematically disadvantage students from equity-deserving backgrounds. The shifts we have outlined, toward distributed understandings of knowledge and agentic approaches to assessment, offer more than theoretical reframing. They provide a foundation for rethinking everyday assessment decisions: what we ask students to do, how we ask them to do it, and who gets to define what counts as having done it well. In a moment where GenAI has destabilised familiar assessment practices, educators have an opportunity to rebuild in ways that recognise diverse ways of knowing and position students as partners in their own learning.

Discussion

Practical Implications

What might educators do differently as a result of this analysis? A starting point could involve critically examining where individualistic assumptions remain embedded in current assessment designs. This might include auditing existing assessments to identify where students are required to demonstrate knowledge in isolation, and exploring alternatives that permit resource use, collaboration, and the supported integration of GenAI, where educators model effective human-AI collaboration rather than prohibiting it (Tai et al., 2023). Where group work already exists, assessment criteria might shift away from tracking individual contributions toward evaluating how teams synthesised perspectives and documented their collaborative processes.

Educators might also experiment with providing students greater agency over assessment parameters. This need not require wholesale redesign; even modest changes, such as offering choice in format, involving students in co-constructing success criteria, or incorporating structured self-assessment, can begin to position students as participants rather than subjects of evaluation (Nieminen et al., 2025). Such approaches may prove particularly valuable for students whose strengths do not align with standardised formats.

For those willing to engage more substantially with grading reform, piloting competency-based or pass/fail approaches within individual units offers a lower-risk entry point. Here, the focus might shift from numerical scores toward documentation of learning processes and the development of navigational capabilities (Bearman et al., 2020). Throughout any such changes, the guiding question remains that posed by Ellis and Lodge (2024): how can we detect learning rather than detecting cheating?

We recognise, however, that these suggestions operate within institutional contexts that may constrain their implementation. Grading policies, professional accreditation requirements, learning management systems, and workload models are often designed around traditional assessment structures. Educators may face resistance from colleagues, students accustomed to familiar formats, or external stakeholders who expect conventional credentialing. These constraints are real and should not be underestimated. Yet institutional change often begins with localised experimentation, and even modest shifts within individual units can generate evidence and momentum for broader transformation.

Theoretical Implications

This paper contributes to assessment scholarship by bringing together two bodies of literature, distributed cognition and student agency, that have largely developed in parallel and examining their combined implications for how success and failure are conceptualised in higher education. In doing so, we extend McDermott and Varenne's (1995) analysis of educational failure as culturally produced, applying their insights specifically to the mechanisms of university assessment in an age of GenAI.

Our analysis also builds on Nieminen's (2024) identification of the paradox of inclusive assessment by suggesting pathways, albeit partial and contested, through which this paradox might be navigated. Rather than proposing that inclusion can be achieved within existing assessment structures, we argue that the structures themselves require fundamental

reconsideration. This positions our work alongside scholars who have called for validity-centred approaches to assessment (Dawson et al., 2024) while extending such arguments explicitly toward equity concerns.

Finally, this paper responds to calls for assessment scholarship to engage more substantively with how GenAI disrupts foundational assumptions about authorship and knowledge creation (Bearman & Ajjawi, 2023). We suggest that GenAI's challenge to individualistic assessment is not merely practical but epistemological, rendering visible the distributed nature of knowing that was always present but more easily ignored in pre-digital contexts (Fawns, 2022). Future scholarship might empirically examine the equity implications of the approaches we have outlined, attending carefully to how distributed and agentic assessment practices play out differently for diverse student populations. Such research could include comparative studies of student outcomes under traditional versus competency-based grading, qualitative investigations of how students from equity-deserving backgrounds experience agentic assessment designs, or action research examining educator attempts to integrate supported GenAI use within collaborative assessment tasks.

Conclusion

In this paper, we have made the case that GenAI presents higher education with a rare opportunity to fundamentally reimagine assessment practices through an equity lens. Specifically, we discuss how the networked and personalised nature of our evolving digital work may lead to greater recognition of the benefits of distributed and agentic approaches to assessment, which could, in turn, better serve our increasingly diverse university student cohort.

However, we must also acknowledge that the transformations we propose are not without their own risks. Distributed assessment, for example, could inadvertently create new barriers for students who struggle in resource-rich, dynamic, and highly interpersonally connected environments. Similarly, agentic assessment, while empowering for some, could also disadvantage students who benefit from clear structure and predetermined expectations, particularly if that is what they have been conditioned to expect. These potential limitations underscore the importance for future research to take care when implementing change and further unpack the nuances of what is suggested here.

We further recognise that our conceptual discussion above is framed to address how GenAI might support more inclusive higher education assessment, with limited attention to broader equity concerns of AI systems. Critical scholarship on AI bias, potential labour displacement, and the concentration of technological power in private entities raises important questions about how these systems might amplify existing inequities in our societies (Bartoletti, 2020; Clifton et al., 2020). It is precisely through such challenging work that we may finally create assessment practices worthy of students' diverse talents and ways of knowing.

Acknowledgements

copywriting of this research as per Crawford et al. (2023). Specifically, Claude (Anthropic, Claude 3.5 Sonnet) was used for proofreading and checking spelling and grammar. The authors confirm that they have met the ethical standards expected as per Purvis and Crawford (2024). The authors list the following CRediT contributions: **Dollinger:** Conceptualisation, Writing – original draft, Writing – review & editing. **Nieminen:** Conceptualisation, Writing – original draft, Writing – review & editing.

References

Adie, L. E., Willis, J., & Van der Kleij, F. M. (2018). Diverse perspectives on student agency in classroom assessment. *The Australian Educational Researcher*, *45*(1), 1-12. https://doi.org/10.1007/s13384-018-0262-2

Ajjawi, R., & Boud, D. (2024). Changing representations of student achievement: The need for innovation. *Innovations in Education and Teaching International*, *61*(3), 597-607. https://doi.org/10.1080/14703297.2023.2192513

Ajjawi, R., Fischer, J., Tai, J., Bearman, M., & Jorre de St Jorre, T. (2023). "Attending lectures in your pyjamas": Student agency in constrained circumstances. *Higher Education*, *86*(6), 1363-1376. https://doi.org/10.1007/s10734-022-00976-9

Ange, B., Wood, E. A., Thomas, A., & Wallach, P. M. (2018). Differences in medical students' academic performance between a pass/fail and tiered grading system. *Southern Medical Journal*, *111*(11), 683-687. https://doi.org/10.14423/SMJ.0000000000000884

Baartman, L. K., & Quinlan, K. M. (2024). Assessment and feedback in higher education reimagined: using programmatic assessment to transform higher education. *Perspectives: Policy and Practice in Higher Education*, *28*(2), 57-67. https://doi.org/10.1080/13603108.2023.2283118

Bartoletti, I. (2020). *An artificial revolution: on power, politics and AI*. Black Spot Books.

Bassett, M. A., Bradshaw, W., Bornsztejn, H., Hogg, A., Murdoch, K., Pearce, B., Webber, C. (2025, October 10). Heads we win, tails you lose: AI detectors in education. *EdArXiv*. https://doi.org/10.35542/osf.io/93w6j_v1

Bearman, M., & Ajjawi, R. (2023). Learning to work with the black box: Pedagogy for a world with artificial intelligence. *British Journal of Educational Technology*, *54*(5), 1160-1173. https://doi.org/10.1111/bjet.13337

Bearman, M., Boud, D., & Ajjawi, R. (2020). New directions for assessment in a digital world. In *Re-imagining university assessment in a digital world* (pp. 7-18). Springer. https://doi.org/10.1007/978-3-030-41956-1_2

Bearman, M., Nieminen, J. H., & Ajjawi, R. (2023). Designing assessment in a digital world: An organising framework. *Assessment & Evaluation in Higher Education*, *48*(3), 291-304. https://doi.org/10.1080/02602938.2022.2069674

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*(6), 893-905. https://doi.org/10.1080/02602938.2024.2335321

Bellah, R., Madsen, R., Sullivan, W. M., Swidler, A., & Tipton, S. M. (1985). *Habits of the heart: Individualism and commitment in American life*. University of California Press.

Bong, W. K., & Chen, W. (2024). Increasing faculty's competence in digital accessibility for inclusive education: A systematic literature review. *International Journal of Inclusive Education*, *28*(2), 197-213. https://doi.org/10.1080/13603116.2021.1937344

Cachia, M., Lynam, S., & Stock, R. (2018). Academic success: Is it just about the grades?. *Higher Education Pedagogies*, *3*(1), 434-439. https://doi.org/10.1080/23752696.2018.1462096

Clifton, J., Glasmeier, A., & Gray, M. (2020). When machines think for us: the consequences for work and place. *Cambridge Journal of Regions, Economy and Society*, *13*(1), 3-23. https://doi.org/10.1093/cjres/rsaa004

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education*, *50*(8), 1397-1410. https://doi.org/10.1080/02602938.2025.2503964

Crawford, J., Cowling, M., & Allen, K. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice, 20*(3). https://doi.org/10.53761/1.20.3.02

Crawford, J., Vallis, C., Yang, J., Fitzgerald, R., & O'Dea, C. (Eds.). (2023). Artificial Intelligence is awesome, but good teaching should always come first. *Journal of University Teaching and Learning Practice*, *20*(7), 1-12. https://doi.org/10.53761/1.20.7.01

Dawson, P. (2020). *Defending assessment security in a digital world: Preventing e-cheating and supporting academic integrity in higher education*. Routledge. https://doi.org/10.4324/9780429324178

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education*, *49*(7), 1005-1016. https://doi.org/10.1080/02602938.2024.2386662

Dollinger, M., Ajjawi, R., Finneran, R., & O'Shea, S. (2024). Conceptualising responsibility and hostility within work-integrated learning placements for students with disabilities. *Disability & Society*, *39*(9), 2217-2235. https://doi.org/10.1080/09687599.2023.2209276

Dollinger, M., Crawford, N., Ajjawi, R., Bearman, M., & Tai, J. (2025). Caught in a loop? Investigating the interplay between time and emotions for university students. *Teaching in Higher Education*, *30*(2), 213-225. https://doi.org/10.1080/13562517.2025.2449642

Elkhoury, E., Ali, A., & Sutherland-Harris, R. (2023). Exploring Faculty Mindsets in Equity-Oriented Assessment. *Journal of University Teaching & Learning Practice, 20*(5). https://doi.org/10.53761/1.20.5.13

Ellis, C., & Lodge, J. (2024, July 9). Stop looking for evidence of cheating with AI and start looking for evidence of learning [LinkedIn post]. LinkedIn. https://www.linkedin.com/pulse/stop-looking-evidence-cheating-ai-start-learning-cath-ellis-h0zzc/

Fawns, T. (2022). An entangled pedagogy: Looking beyond the pedagogy—technology dichotomy. *Postdigital Science and Education*, *4*(3), 711-728. https://doi.org/10.1007/s42438-022-00302-7

Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. *Computers and Education: Artificial Intelligence*, *6*, 100250. https://doi.org/10.1016/j.caeai.2024.100250

French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, *88*(3), 893-918. https://doi.org/10.1007/s10734-023-01148-z

Gibson, P., Clarkson, R., & Scott, M. (2022). Promoting potential through purposeful inclusive assessment for distance learners. *Distance Education*, *43*(4), 543-555. https://doi.org/10.1080/01587919.2022.2143321

Guenther, J., Fuqua, M., Ledger, S., Davie, S., Cuervo, H., Lasselle, L., & Downes, N. (2023). The perennials and trends of rural education: Discourses that shape research and practice. *Australian and International Journal of Rural Education*, *33*(3), 1-29. https://search.informit.org/doi/abs/10.3316/informit.459645414214719

Herdt, J. A. (2019). *Forming humanity: Redeeming the German Bildung tradition*. University of Chicago Press. https://www.degruyterbrill.com/document/doi/10.7208/9780226618517/html

Hiner, N. R. (1973). The cultural function of grading. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *47*(6), 356-361. https://doi.org/10.1080/00098655.1973.11477761

Jorre de St Jorre, T., Boud, D., & Johnson, E. D. (2021). Assessment for distinctiveness: Recognising diversity of accomplishments. *Studies in Higher Education*, *46*(7), 1371-1382. https://doi.org/10.1080/03075079.2019.1689385

Kjærgaard, A., Mikkelsen, E. N., & Buhl-Wiggers, J. (2023). The gradeless paradox: Emancipatory promises but ambivalent effects of gradeless learning in business and management education. *Management Learning*, *54*(4), 556-575. https://doi.org/10.1177/13505076221101146

Kilpatrick, S., Jones, T., & Barrett, M. (2003). *Defining learning communities*. Launceston, Tasmania: Centre for Research and Learning in Regional Australia. https://www.aare.edu.au/data/publications/2003/jon03441.pdf

La Chimea, T., Kanji, Z., & Schmitz, S. (2020). Assessment of clinical competence in competency-based education. *Canadian Journal of Dental Hygiene*, *54*(2), 83-92. https://pmc.ncbi.nlm.nih.gov/articles/PMC7668267/

Love, J., & Daniel, J. L. (2025). Ungrading: Socially just assessment practices for the PA classroom. *Journal of Social Equity and Public Administration*, *3*(1), 173-196. https://doi.org/10.24926/jsepa.v3i1.5891

McArthur, J. (2016). Assessment for social justice: The role of assessment in achieving social justice. *Assessment & Evaluation in Higher Education*, *41*(7), 967-981. https://doi.org/10.1080/02602938.2015.1053429

McDermott, R., & Varenne, H. (1995). Culture as disability. *Anthropology & Education Quarterly*, *26*(3), 324-348. https://doi.org/10.1525/aeq.1995.26.3.05x0936z

McInnis, C. (2001). Researching the First Year Experience: Where to from Here? *Higher Education Research & Development*, *20*(2), 105-144. https://doi.org/10.1080/07294360125188

Naylor, R., & Mifsud, N. (2020). Towards a structural inequality framework for student retention and success. *Higher Education Research & Development*, *39*(2), 259-272. https://doi.org/10.1080/07294360.2019.1670143

Nieminen, J. H., Yan, Z., & Boud, D. (2025). Self-assessment design in a digital world: Centring student agency. *Assessment & Evaluation in Higher Education*, *50*(2), 323-337. https://doi.org/10.1080/02602938.2025.2467647

Nieminen, J. H. (2023). Unveiling ableism and disablism in assessment: A critical analysis of disabled students' experiences of assessment and assessment accommodations. *Higher Education*, *85*(3), 613-636. https://doi.org/10.1007/s10734-022-00857-1

Nieminen, J. H. (2024). The paradox of inclusive assessment. *Assessment & Evaluation in Higher Education*, *49*(8), 1358-1370. https://doi.org/10.1080/02602938.2024.2419604

Nieminen, J. H., Dollinger, M., & Finneran, R. (2025). 'There was very little room for me to be me': The lived tensions between assessment standardisation and student diversity. *Assessment & Evaluation in Higher Education*, *50*(2), 308-322. https://doi.org/10.1080/02602938.2024.2388699

Nieminen, J. H., Dollinger, M., & Zou, T. X. (2025). Student partnership in assessment: What works, for whom, why and under what circumstances? *Review of Education, (13)*2. https://doi.org/10.1002/rev3.70059

Nulty, D. D. (2011). Peer and self-assessment in the first year of university. *Assessment & Evaluation in Higher Education*, *36*(5), 493-507. https://doi.org/10.1080/02602930903540983

Ogunleye, B., Zakariyyah, K. I., Ajao, O., Olayinka, O., & Sharma, H. (2024). Higher education assessment practice in the era of generative AI tools. *arXiv preprint* arXiv:2404.01036. https://doi.org/10.48550/arXiv.2404.01036

O'Shea, S., & Delahunty, J. (2018). Getting through the day and still having a smile on my face! How do students define success in the university learning environment?. *Higher Education Research & Development*, *37*(5), 1062-1075. https://doi.org/10.1080/07294360.2018.1463973

O'Shea, S., Lysaght, P., Roberts, J., & Harwood, V. (2016). Shifting the blame in higher education–social inclusion and deficit discourses. *Higher Education Research & Development*, *35*(2), 322-336. https://doi.org/10.1080/07294360.2015.1087388

Pang, T. Y., Kootsookos, A., & Cheng, C. T. (2024). Artificial intelligence use in feedback: A qualitative analysis. *Journal of University Teaching and Learning Practice*, *21*(6), 108-125. https://doi.org/10.53761/40wmcj98

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education*, *36*(8), 879–896. https://doi.org/10.1080/03075079.2010.483513

Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency and learning analytics: An exploration. *Journal of Learning Analytics*, *3*(1), 159-182. https://doi.org/10.18608/jla.2016.31.10

Rae, M., Tai, J., & Dawson, P. (2025). Giving voice to women students: designing oral assessments for inclusion and validity. *Assessment & Evaluation in Higher Education*, 1-14. https://doi.org/10.1080/02602938.2025.2580622

Rapchak, M., Hands, A. S., & Hensley, M. K. (2023). Moving toward equity: Experiences with ungrading. *Journal of Education for Library and Information Science*, *64*(1), 89-98. https://doi.org/10.3138/jelis-2022-0017

Rasooli, A., Pitt, E., Turner, J., Varga-Atkins, T., Moindrot, W., Preston, S., & Asgari, S. (2025). Understanding students' perceptions of fairness in group work assessment: Key conditions and implications. *Assessment & Evaluation in Higher Education*, *50*(2), 292-307. https://doi.org/10.1080/02602938.2025.2502587

Rasul, T., Nair, S., Kalendra, D., Balaji, M. S., de Oliveira Santini, F., Ladeira, W. J., Islam, J. U., & Hossain, M. U. (2024). Enhancing academic integrity among students in GenAI Era: A

holistic framework. *The International Journal of Management Education*, *22*(3), 101041. https://doi.org/10.1016/j.ijme.2024.101041

Ryan, A., & Judd, T. (2022). From traditional to programmatic assessment in three (not so) easy steps. *Education Sciences*, *12*(7), 487. https://doi.org/10.3390/educsci12070487

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, *11*(1), 1-31. https://www.tandfonline.com/doi/abs/10.1207/s15326977ea1101_1

Schuwirth, L. W., & Van der Vleuten, C. P. (2011). Programmatic assessment: from assessment of learning to assessment for learning. *Medical teacher*, *33*(6), 478-485. https://doi.org/10.3109/0142159X.2011.565828

Shen, Y., Zhu, Y., & Yang, R. (2025). Integrating traditional knowledge of non-Western humanities and social sciences: a research agenda through the Chinese lens. *Studies in Higher Education*, *50*(9), 1951-1965. https://doi.org/10.1080/03075079.2024.2407051

Smith, D. P., Sokoya, D., Moore, S., Okonkwo, C., Boyd, C., Lacey, M. M., & Francis, N. J. (2025). Embedding Generative AI as a digital capability into a year-long skills program. Journal of University Teaching and Learning Practice, 22 (3). https://doi.org/10.53761/fh6q4v89

Spitzer, M. W. H., Langsdorf, L. E., Richter, E., & Schubert, T. (2025). Low-performing students benefit mostly from open-book examinations. *Computers and Education Open*, *8*, 100239. https://doi.org/10.1016/j.caeo.2024.100239

Spring, L., Robillard, D., Gehlbach, L., & Moore Simas, T. A. (2011). Impact of pass/fail grading on medical students' well-being and academic outcomes. *Medical Education*, *45*(9), 867-877. https://doi.org/10.1111/j.1365-2923.2011.03989.x

Steele, C., Gower, G., & Bogachenko, T. (2024). Creating and enacting culturally responsive assessment for First Nations students in higher education settings. *Australian Journal of Education*, *68*(2), 84-102. https://doi.org/10.1177/00049441241258496

Stoesz, B. M. (2023). Lies, lies, lies: Detecting deception and implications for investigations of academic cheating. In *Academic integrity in the social sciences: Perspectives on pedagogy and practice* (pp. 53-68). Springer. https://doi.org/10.1007/978-3-031-43292-7_4

Stommel, J. (2023). *Undoing the grade: Why we grade, and how to stop*. Denver: Hybrid Pedagogy, Incorporated.

Stone, C., Freeman, E., Dyment, J. E., Muir, T., & Milthorpe, N. (2019). Equal or equitable?: The role of flexibility within online education. *Australian and International Journal of Rural Education*, *29*(2), 26-40. https://search.informit.org/doi/abs/10.3316/informit.639619709196818

Tai, J., Ajjawi, R., Bearman, M., Boud, D., Dawson, P., & Jorre de St Jorre, T. (2023). Assessment for inclusion: Rethinking contemporary strategies in assessment design. *Higher Education Research & Development*, *42*(2), 483-497. https://doi.org/10.1080/07294360.2022.2057451

Taylor, L. (2021). There is an app for that: Technological solutionism as COVID-19 policy in the global north. In *The new common: How the COVID-19 pandemic is transforming society* (pp. 209-215). Springer. https://doi.org/10.1007/978-3-030-65355-2_13

Walden, P. R. (2022). Student motivation, anxiety and pass/fail grading: A SoTL project. *Teaching and Learning in Communication Sciences & Disorders*, *6*(1), 14. https://doi.org/10.30707/TLCSD6.1.1649037808.651639

Zarb, M., McDermott, R., Martin, K., Young, T., & McGowan, J. (2023, October). Evaluating a pass/fail grading model in first year undergraduate computing. In *2023 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE. https://doi.org/10.1109/FIE58773.2023.10343276