



## Assessment after Artificial Intelligence: The Research We Should Be Doing

Dr Thomas Corbin<sup>a</sup>, Professor Margaret Bearman<sup>a</sup>, Professor David Boud<sup>a</sup>, Dr Nicole Crawford<sup>a</sup>, Professor Phillip Dawson<sup>a</sup>, Associate Professor Tim Fawns<sup>b</sup>, Professor Michael Henderson<sup>b</sup>, Professor Jason Lodge<sup>c</sup>, Assistant Professor Jiahui Jess Luo<sup>d</sup>, Professor Kelly Matthews<sup>c</sup>, Associate Professor Kelli Nicola-Richmond<sup>a</sup>, Dr Juuso Henrik Nieminen<sup>a</sup>, Associate Professor Nicole Pepperell<sup>e</sup>, Dr Zachari Swiecki<sup>b</sup>, Associate Professor Joanna Tai<sup>a</sup>, Dr Jack Walton<sup>a</sup>

<sup>a</sup> Deakin University, Australia; <sup>b</sup> Monash University, Australia; <sup>c</sup> University of Queensland, Australia; <sup>d</sup> The Education University of Hong Kong, Hong Kong; <sup>e</sup> University of Technology Sydney, Sydney

### Abstract

The emergence of widely available artificial intelligence (AI) tools has made assessment in higher education increasingly uncertain. Familiar (if problematic) assumptions about what assessment does or should measure, who or what is being assessed, and how judgments are made are all being re-examined. Educators and researchers are experimenting with new assessment designs, but the emerging research landscape is fragmented and difficult to navigate. There is little shared sense of what kinds of studies are most needed or how their findings might connect. To address this, a group of leading assessment scholars met in Melbourne, Australia in September of 2025 to develop a collective research agenda to help guide and connect future inquiry. This paper presents the outcomes of that collaboration, a set of guiding principles and framing questions – why, who, what, how, and where we assess – that together offer a structure for guiding and supporting the research we should be doing on assessment after AI.

### Editors

Section: Editor-in-Chief  
Editor-in-Chief: Dr Joseph Crawford

### Publication

Submission: 5 November 2025  
Revised: 18 November 2025  
Accepted: 27 November 2025  
Published: 4 December 2025

**Copyright** © by the authors, in its year of first publication. This publication is an open access publication under the Creative Commons Attribution [CC BY-ND 4.0](#) license.

### Practitioner Notes

1. Not all AI use is the same, so assessment research and policy need to say exactly what students are doing with AI, and in what context.
2. Assessment research and policy should build on what we already know about how students learn, write, and share labour, because these practices did not suddenly begin with generative AI.
3. Any changes to assessment must consider equity, access, and fairness.
4. We need empirical and conceptual research that is careful, connected, and useful, so it can actually help teachers and institutions make defensible assessment decisions in a time of AI.

### Keywords

Assessment, AI-integrated assessment, artificial intelligence, programmatic assessment, assessment integrity.

### Citation:

Corbin, T., Bearman, M., Boud, D., Crawford, N., Dawson, P., Fawns, T., Henderson, M., Lodge, J., Luo, J., Matthews, K., Nicola-Richmond, K., Nieminen, J., Pepperell, N., Swiecki, Z., Tai, J., & Walton, J. (2025). Assessment after Artificial Intelligence: The Research We Should Be Doing. *Journal of University Teaching and Learning Practice*, 22(7). <https://doi.org/10.53761/w3x5y804>

## Introduction

Teachers, students, and universities are all struggling to adapt to the emergence of artificial intelligence (AI), and nowhere is this more evident than in assessment. Each semester brings uncertainty about what a student's work really shows, how much and what type of AI use should be acceptable, and whether familiar tasks will still provide meaningful evidence. Both teachers (Corbin et al., 2025a; Farazouli et al., 2025) and students (Bearman et al., 2025; Walton et al., 2025) are left to draw boundaries regarding acceptable AI use which they feel they cannot defend with confidence, while institutions face the prospect that if assessment is no longer trusted, neither are their degrees. Of course, assessment was not perfect before AI became commonly accessible, and it has always been entwined with complex issues related to the promotion, sustainability, and assurance of learning (Boud, 2000). However, whilst AI certainly poses some new problems and it renews the significance of longer-standing issues, it may simultaneously provide a lever to make some changes that have been well-overdue. With all this in mind, it is little wonder that teachers and teaching institutions are struggling.

Many initial responses to the challenges posed by AI to assessment promised direct solutions, premised on the idea of 'containing' disruptions to assessment. Detection software is one such example, promoted as a way to identify AI-generated text and thereby contain the disruption, allowing business as usual to continue. That promise has not held. Detection systems are inconsistent, easily outmanoeuvred, and in some cases discriminatory (Ardito, 2025; Bassett et al., 2025). More fundamentally, the idea of containment misdescribes the problem, treating AI as an isolated threat rather than a shift in the conditions of assessment itself.

We are increasingly starting to realise that, together, AI and assessment form a wicked problem (Corbin et al., 2025b). There is no 'correct' solution waiting to be discovered, only better or worse responses where every decision comes with lived consequences. Decisions to design tasks to reduce AI use can narrow what students are able to do or demonstrate; shifting to more open formats where students are free to use AI makes it harder to judge students' capability in relation to what is being assessed. Each move involves trade-offs that affect teachers, students, and institutions. Teachers, however, need to make decisions. They have students in class today and cannot put off decision making until a (likely non-existent) perfect solution is found. These decisions are hard not least because it is often unclear what is actually at stake, what is really being traded off and what might be possible to gain. It is at this moment of uncertainty that good research becomes especially important, to shine light on the unknowns and help students, teachers, learning designers, higher education leaders, policy makers and others to make informed decisions.

When widely-available generative AI first emerged, conceptual and empirical resources to guide practice were scarce, so the sector needed to act upon its best conjectures on what needed to be done, based on what had come before (e.g. Lodge et al. 2023). Now, the landscape of generative AI tools and use is becoming clearer and research on assessment and AI is expanding rapidly. Nonetheless, the sector does appear to be struggling to accumulate useful insights which advance knowledge and practice. Despite the volume of work published across conventional and emerging outlets, the field has not yet made the kind of headway we might expect from such collective effort. We suggest that an orienting research agenda might help.

The research agenda we propose in this paper emerged from a three-day meeting in Melbourne in September 2025 held in the Deakin University Centre for Research in Assessment and Digital Learning (CRADLE). This event brought together local and international scholars working in assessment and adjacent areas of higher education. These scholars, who became the authors of this agenda, included both established and emerging researchers whose recent work had already been shaping debates about assessment and AI. The results, which we share here, are not a record of complete consensus but the distillation of many productive disagreements, offered in the spirit of helping the field move forward through more connected and cumulative inquiry.

In what follows, we first focus on a set of guiding principles that we suggest might underpin future research. Second, we offer more specific items that might belong on such an agenda. We offer this agenda as a living discussion paper, as a shared starting point of suggestive orientation rather than a prescription. Our intention is not to “set the agenda” that others must follow, but rather to suggest a vision of needed research based on our own understanding and experience as a place for dialogue and refinement as researchers and practitioners alike grapple with how assessment can remain meaningful, legitimate, and educative in a time of AI.

Throughout this piece, we adopt the broader term “AI” to encompass the wider spectrum of AI technologies that are influencing higher education assessment. While generative AI – a subset of AI capable of generating a variety of content – has been the primary driver of many recent debates described earlier, its rapid development has also brought attention to other forms of AI that, though not generative in nature, are becoming increasingly relevant and widely used in higher education assessment (Swiecki et al., 2022). The proposed agenda has been shaped significantly by the rise of generative AI but is also grounded in the broader landscape where various types of existing and emerging AI technologies are transforming assessment practices in higher education (e.g., predictive debugging tools and machine translation).

## **Principles of Artificial Intelligence Research**

We believe that what makes an agenda more than merely a list of important topics is the orientation it gives to research. That orientation is especially important in the current moment since the arrival of AI in widely accessible forms (such as ChatGPT subscriptions) has unsettled assessment in ways that make research (even on important topics) unusually vulnerable to shallow framing, premature claims, and false ‘silver bullet’ fixes. An agenda may identify where inquiry is needed, but it may struggle to help ensure that the research done in those spaces will be worth having. It is important to identify and share new priorities, but it is perhaps more important to ensure that the work within those areas accumulates into something coherent, relevant, and trustworthy. For these reasons, this agenda begins with the following set of guiding principles:

### **Principle 1. Research should clarify the multiple contextualised meanings of assessment**

While the emergence of AI has prompted the creation of this agenda, the question we address is not primarily about AI but rather about assessment itself. Assessment, however, has never served a single purpose (Boud, 1995; Broadfoot, 1996). It assures the quality of learning outcomes, shapes what and how students study, provides feedback for improvement, prompts students’ self-judgements and communicates achievement to the wider world (Boud & Soler, 2016). These purposes frequently coexist in tension, and the current disruption renews the visibility of these tensions.

Most immediate concerns have centred on issues of assurance – namely, whether grades and degree outcomes can still be trusted to represent student capability (Lodge et al., 2025). While that concern is justified, it risks taking for granted pre-existing issues related to grading (e.g. Sadler, 2013) and the representation of achievement (e.g. Jorre de St Jorre et al., 2021), and the overwhelming attention being given to this concern risks eclipsing assessment's other functions. If research focuses only on detecting or containing AI use, or on how learning outcomes can be assured where students can use AI, it risks narrowing the field to the mechanics of verification rather than the broader educational purposes that assessment serves (Boud, 2000). Assessment research, then, we suggest, must begin with renewed clarity about these purposes. It must ask which aspects of assessment are non-negotiable and which can be adapted, what forms of evidence can still act as evidence of learning, and how changes in practice affect the learning and recognition of students.

### **Principle 2. Research should clarify the multiple contextualised meanings of AI**

The second principle that we suggest should guide research on assessment after AI, is that AI should not be treated as a monolith. Current debates often refer to "AI" or "GenAI" as if these were singular entities with uncontroversial definitions. In reality, however, these terms bundle together unique, distinct practices which are often not reducible to a shared title (Bearman & Ajjawi, 2024). These are practices such as summarising, paraphrasing, generating drafts, simulating dialogue, or planning tasks. Each of these practices involves distinct kinds of labour, poses different challenges for assessment, and carries different implications for learning. In the context of research, this means examining and describing the specific practices at play: what people do with AI, in what contexts, and to what ends. This issue is likely to become increasingly important as AI is built into everyday items which are used or even relied on for legitimate purposes (AI enabled prescription reading glasses, for example). Clarifying AI use in this way allows research to produce findings that are interpretable, comparable, and relevant to assessment design.

The most immediately concerning practical consequence of not doing so, of treating AI monolithically, is that research becomes difficult to use. When a study reports that "students used AI", it provides little guidance for practitioners trying to make decisions or for researchers attempting to build on the findings. A teacher reading such research cannot determine whether the findings apply to their students using AI to check grammar, to generate essay outlines, or to simulate dialogue with historical figures. A researcher cannot determine whether apparent contradictions between studies reflect genuine differences in educational outcomes or simply differences in what "AI use" meant in each context. For research to be interpretable and actionable, the specific practices must be described with enough precision that others can recognise whether they are studying, teaching, or regulating comparable phenomena.

### **Principle 3. Research should build on continuities**

The fact that generative AI tools only became widely available in late 2022 makes it tempting to treat that moment in time as the cut off for relevant research on AI and assessment. It is very easy to say that the technologies are new, the disruption is sudden, and the problems are unique. However, saying that risks cutting current debates off from the longer histories of both the relevant theories (of learning, education, and technology) as well as the practices at stake. This tendency

risks amounting to a kind of historical amnesia, forgetting, for example, that summarising, paraphrasing, delegating intellectual labour, and generating text are not novel activities but long-standing features of reading, writing, and feedback. Once we recognise that the real object of study is not “AI” itself but the practices it mediates, those histories become indispensable.

The risk of ignoring continuities is not merely that we fail to cite relevant prior work, but that we produce research that is conceptually impoverished. When researchers treat AI-mediated practices as entirely novel, they often develop ad hoc explanatory frameworks that are less sophisticated than the theoretical resources already available. We see studies rediscovering insights about collaboration, delegation, or textual borrowing that have been well understood in other contexts, now presented as discoveries about AI. We see frameworks for understanding learning or capability that lack the conceptual depth and precision of earlier theories, distinguished only by their attention to AI as a variable. We also risk abandoning valuable assessment tasks that appear vulnerable to AI without understanding what we lose in doing so or investigating whether that loss is necessary. The essay perhaps best illustrates this tension, as while AI can generate essay-like text, the essay's value lies in its capacity to develop exploratory thinking, self-regulated learning, and engagement with uncertainty, each being capacities that may warrant preservation through thoughtful redesign rather than wholesale abandonment (Corbin et al., 2025c).

The result of failing to build on continuities is the risk of building a body of research that appears to grow rapidly but in fact advances understanding slowly, where new studies rebuild explanatory machinery from scratch rather than refining and extending what we already know about the practices that now includes AI to some degree. Therefore, we suggest that good research on AI and assessment should always aim to build on appropriate continuities that travel through the knowledge accumulated prior to the emergence of AI. Doing so, we suggest, is likely to produce research that is deeper, more cumulative, and more useful to practice than research which does not engage with the long tail of our knowledge.

#### **Principle 4. Research should consider inclusion, equity, ethics and social justice**

We see inclusion, equity, ethics and social justice as necessary guiding principles for future research. We do not wish these ideas to emerge as one ‘theme’ or ‘topic’, but instead, we propose that any research initiative must see these values as something fundamental (McArthur, 2018). For example, research on AI in assessment must acknowledge the issues of algorithmic justice and potential biases with AI, the accessibility of AI technologies, data sovereignty, and the potentially shifting landscapes of ethics and trust in higher education (see McDermott & Eaton, 2025). Likewise, research on assessment after AI should build on ethical and self-reflexive foundations, aiming to produce work that considers the multiple voices of the sector and helps higher education institutions to strive towards social good.

Together, these four principles are intended to shape not only what we study but how we approach that study. These principles do not dictate specific methods or topics, but they do establish a foundation for research that is contextually sensitive, historically grounded, and oriented toward cumulative insight. With these principles as a foundation, we now turn to the specific domains where we believe research is most urgently needed.

## **A Research Agenda**

If the guiding research principles outlined above orient how research on AI and assessment should be pursued, the next step is to identify where that research is most needed. The purpose of this agenda is not to predict every direction the field will take, but to map out domains of inquiry that we believe demand urgent and sustained attention.

We have organised this agenda around what we believe to be longstanding questions fundamental to assessment research - why, who, what, how, where, and what if. We have done this to provide a shared framework that can help connect what might otherwise appear to be disconnected lines of inquiry. These questions deliberately overlap and nest within each other; research on how to design a task cannot be separated from questions about what is being assessed or why. This overlap is intentional (Purvis et al., 2024). Many studies may already be addressing aspects of these questions without explicitly framing them as such. Our hope is that by offering this structure, researchers can recognise how their work connects to broader concerns about assessment's legitimacy, purpose, and practice. This framing may help researchers situate their contributions within larger conversations, making it easier for the field to build cumulative knowledge rather than producing isolated findings. The questions are not meant to partition the research space into separate domains, but to illuminate the interconnections that can make individual studies more interpretable and collectively more powerful. We will begin with the most fundamental question, why do we assess in the first place?

### **Why do we assess?**

The first agenda item of why points towards the need for research around the justification of assessment. Assessment has never been a neutral measurement act. Why we assess concerns not merely the efficiency or reliability of judgments, but the social purposes that justify them – fairness, legitimacy, and the credibility of qualifications in the public sphere. These purposes demand renewed attention in research. When authorship and capability become increasingly opaque, the originally tenuous connection between what is produced and what it represents follows suit.

Without research into why we assess, our responses to the AI challenge risk collapsing into piecemeal technical fixes that preserve measurement routines while eroding the moral and social grounds on which those routines are legitimated. Why questions in assessment research invite inquiry into how judgments retain credibility, how credentials maintain public trust, and how fairness is re-interpreted when human and machine contributions intertwine. They ask what kinds of value assessment now defends and how those values are expressed through institutional practice, professional standards, and the expectations of the wider world.

Research is therefore needed in order explore how assessment can continue to justify itself in both educational and civic contexts. How can assessment uphold equitable treatment, sustain the moral legitimacy of credentials, and maintain confidence in the relationship between learning and its evidence? Research on why we assess enables educators, institutions, and researchers to determine what should be preserved, what should be reimaged, and what forms of judgment can still be defended as legitimate in a time when the conditions of authorship and evidence have changed.

## **Who should be involved in assessment?**

The second agenda domain of who concerns the people and institutions whose participation gives assessment its legitimacy. Assessment has never been a simple exchange between student and teacher. It has always depended on a broader network of educators, designers, peers, professional bodies, employers, and the public who rely on universities to certify capability. Research must examine how this network is being reconfigured; who is involved in designing, judging, and validating assessment, whose recognition confers value, and whose vision of assessment are we supporting in our work?

At its core, the question of who also concerns attribution: whose capability does a grade actually describe (Luo & Dawson, 2025) and under what conditions? Assessment has long been premised on the idea that the student submitting work is the bearer of the capability behind it. That assumption was always partial, since learning naturally draws on the work of others and on social and material resources. In part, AI has made this more visible. When ideas, words, or structures are co-produced with AI systems, the subject of assessment becomes more uncertain. Who or what is represented by the grade? Who claims ownership and responsibility for the work's meaning? These are not new questions, but they add to the urgency regarding systematic investigation of how authorship and accountability are negotiated in hybrid conditions (Cheng et al., 2024).

Research should therefore explore how authority and responsibility are shared across educators, students, designers, employers, and automated systems; which forms of judgment require human interpretation and recognition; and how legitimacy is maintained when assessment outcomes circulate beyond the university. Understanding who is involved in assessment, and on what grounds, will be crucial for ensuring that assessment remains a credible, recognitive, and publicly defensible practice of judgment.

## **What can and should we be assessing?**

The question of what we assess asks what kinds of learning, understanding, or capability assessment now claims to evidence. Every assessment embodies an assumption of purpose and a theory of what matters, and the current disruption exposes how those theories are encoded in learning outcomes. Learning outcomes are likely the pivotal concept, yet the emergence of hybrid human-AI capability unsettles what those outcomes mean and how they can be evidenced.

Research after AI therefore needs to re-examine how learning outcomes are constituted; what forms of knowing, reasoning, and collaboration they presuppose, and whether those forms still capture the kinds of capability that education values. Some of this work is practical: determining how assurance processes and standards might evolve so that they remain defensible while accommodating distributed forms of authorship and judgment. Some of it is philosophical: asking what remains distinctively human in learning and whether such qualities can, or should, be rendered assessable.

Empirical and conceptual inquiry are both needed. Conceptually, to articulate why certain capacities (critical judgment, discernment, responsibility, creativity) continue to matter for the purposes of education; empirically, to identify what kinds of evidence can still track them. Without clarity about what assessment claims to evidence, decisions about design and policy risk becoming arbitrary. Work in this area provides the conceptual backbone for the rest of the agenda,

defining the evolving objects of assessment so that how, who, and where can each be addressed on defensible ground.

### **How and when should we assess?**

The question of how directs attention to the ways assessment practices generate and interpret evidence of learning. It asks what kinds of tasks, interactions, conditions and artefacts can still link students' activity to defensible claims about their capability when AI shapes how work is produced. Assessment has always involved translation (from what learners do, to what teachers see, to what institutions record) but AI complicates each step. Research under this heading therefore explores how assessment can make those processes visible and interpretable. What forms of task design, supervision, feedback, and reflection help connect outcomes to the learning they represent? How can evidence be triangulated across drafts, dialogue, and process data without collapsing into surveillance? And how might AI itself be used to support, rather than obscure, the interpretive work of assessment? Addressing these questions requires methodological imagination: studies that test how different designs perform under real educational conditions, identify where validity breaks down, and develop shared language for evaluating the quality of evidence in hybrid (human and AI) production. The how of assessment is therefore practical as well as conceptual, concerned not only with task mechanics but with maintaining interpretive credibility. In choosing how and when to assess, we must keep in mind long-standing research that identifies that it is not the intentions of the teacher/designer, nor the intrinsic qualities of the task, but the ways in which students interpret and act on what is put before them that lead to a particular influence (Boud, 2000).

### **Where are the appropriate sites of assessment for which purpose?**

The question of where assessment takes place, on what objects it focuses, highlights issues of form and format. At what level, and in what settings for which purposes, should assessment now occur? The traditional classroom or unit is no longer the only or even primary site where assessment is enacted. Students may complete work in digital and AI-rich environments, teachers may design or grade within external tools, and universities may increasingly rely on platform analytics or credentialing systems to confirm achievement. Research under this heading examines how these shifting locations affect the credibility, comparability, and educational value of assessment. Should judgments about capability be anchored in single tasks, in programmatic portfolios, or in longitudinal records of practice? How does the growing involvement of third-party systems redistribute institutional authority and accountability? And how can coordination across sites ensure that assessment remains coherent to students and defensible to the public? These questions are both practical and structural: they concern how universities design assessment ecosystems that can accommodate AI without fragmenting responsibility or losing trust. The where of assessment therefore invites inquiry into governance and infrastructure; the architectures through which assessment occurs, must be verified, and made to matter.

### **What if...?**

The question of what if invites research that re-imagines assessment itself rather than adapting existing forms. Much of the current response to AI has been reactive, seeking to restore stability or contain disruption. Yet assessment has always evolved through moments of conceptual risk. What if research asks generative questions about how assessment might be otherwise: what if



evidence of learning need not take the form of artefacts, or if recognition could occur through collaborative or longitudinal processes across units rather than discrete tasks? Inquiry of this kind does not aim for immediate implementation but for discovery, for testing new assumptions, surfacing neglected possibilities, and articulating futures that current practices cannot yet support. Such work is essential if the field is to remain intellectually alive, ensuring that assessment research not only adapts to AI but also imagines what assessment could become. By articulating these speculative possibilities, ‘what if?’ research also equips institutions with materials for the strategic foresight needed to plan pragmatically for assessment systems that remain robust and educationally defensible amid ongoing technological change

## **Discussion**

These six domains do not exhaust every possible line of inquiry. Rather, they represent the areas we believe research effort is most urgently needed if assessment is to respond to the challenges posed by AI. We highlight them because they capture the key unknowns that teachers and institutions are already weighing as they make real-time decisions: What kinds of assessment designs can still do their work, what exactly is being assessed and for what purpose, how assessment unfolds in practice, and what moral responsibilities and political stakes are at play. Without sustained attention to these uncertainties, research risks remaining piecemeal and disconnected from the struggles of practice. The domains are not prescriptions of method or topic, but invitations, and points of orientation that can help connect otherwise fragmented projects into a shared conversation. Each domain marks a space where research can both clarify practice and advance theory, and together they provide a map for the next stage of scholarship in this field.

Yet identifying where research is needed does not resolve the question of how that research should be conducted. If this agenda is to fulfil its purpose, that of helping teachers and institutions make better-informed decisions about assessment in an AI context, then the research must be able to speak with those it aims to serve. This likely means considering methodologies and research approaches which involve educators, students, assessment designers, institutional leaders, and others who are directly impacted by assessment decisions not merely as research subjects but as active participants in the research process itself (Cook-Sather, 2018; Matthews, 2018; Cook-Sather & Matthews, 2023; Nicola-Richmond et al., 2025). Participatory and co-design approaches can ensure that research questions emerge from the actual dilemmas being faced in practice, that studies are designed to generate actionable insight, and that findings are interpretable and relevant to those who must act on them (Bovill et al., 2021). Such approaches also help to ensure that research attends to the diverse contexts in which assessment occurs, across disciplines, institution types, and student populations, rather than producing findings that reflect only narrow or privileged perspectives (Cook-Sather, 2022).

Building research in partnership with practitioners is not only a matter of inclusion or relevance; rather, it is also useful for creating the kind of sustained connection between inquiry and practice that this moment demands. Research that remains disconnected from assessment practice as it is being navigated by teachers and universities risks arriving too late or speaking to questions practitioners are not interested in. When research is conducted collaboratively with those navigating these challenges in real time, it becomes directly responsive to the decisions being made. It can test emerging practices as they develop, identify what works and what fails in specific contexts, surface unintended consequences as they arise, and refine approaches iteratively

based on lived experience. This creates a stronger relationship between inquiry and action where research informs practice, practice reveals new questions, and those questions shape subsequent research. By building these connections deliberately into how we conduct research, we ensure that inquiry generates not just knowledge about assessment in this time of AI, but usable knowledge for those who must make assessment work in conditions of ongoing uncertainty. In this way, how we conduct research becomes inseparable from the agenda's core aim of building research which works to support practitioners and institutions as they navigate the difficult work of assessing in ways which are meaningful, legitimate, and educative in a time of profound change.

## Conclusion

Teachers and institutions are making decisions about assessment in an AI context today. They are experimenting with designs, revising policies, and negotiating legitimacy in real time as they teach students who look to them for instruction and support. Research is needed to support those decisions and help anticipate their consequences. The six domains outlined here are intended as building blocks for that knowledge, areas where research can contribute most directly to the challenges now being felt in practice. For researchers, we hope that these domains can provide a way of situating projects so that their relevance is immediately clear. Using this language connects local studies to a wider conversation, making them easier to compare and build on.

The future of assessment will be defined by trade-offs rather than simple solutions. Research that takes up these domains can help to clarify what those trade-offs are, and what they mean in different contexts. In this way the agenda we offer is, of course, not an answer in itself but hopefully a useful resource, a way of making visible the choices ahead, and of supporting practice and policy with evidence that accumulates rather than fragments. Others will hopefully refine what we have suggested, challenge it, and extend it in directions we have not anticipated. The questions raised by AI in regard to assessment are larger than any single project or group. What matters is that our combined inquiry moves forward in ways that are cumulative, connected, and attentive to the real dilemmas of practice.

## Acknowledgements

The authors disclose that they have no actual or perceived conflicts of interest. The authors disclose that they have not received any funding for this manuscript beyond resourcing for academic time at their respective university. The authors have not used artificial intelligence in the ideation, design, or write-up. The authors confirm that they have met the ethical standards expected of the Journal. The authors list the following CRediT contributions: **Thomas Corbin** (Project administration, Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Margaret Bearman** (Project administration, Conceptualization, Writing - Original Draft, Writing - Review & Editing), **David Boud** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Nicole Crawford** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Phillip Dawson** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Tim Fawns** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Michael Henderson** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Jason Lodge** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Jiahui Jess Luo** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Kelly Matthews**

(Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Kelli Nicola-Richmond** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Juuso Nieminen** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Nicole Pepperell** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Zachari Swiecki** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Joanna Tai** (Conceptualization, Writing - Original Draft, Writing - Review & Editing), **Jack Walton** (Conceptualization, Writing - Original Draft, Writing - Review & Editing).

## References

- Ardito, C. G. (2025). Generative AI detection in higher education assessments. *New Directions for Teaching and Learning*, 2025, 11–28. <https://doi.org/10.1002/tl.20624>
- Bassett, M. A., Bradshaw, W., Bornsztejn, H., Hogg, A., Murdoch, K., Pearce, B., & Webber, C. (2025). *Heads we win, tails you lose: AI detectors in education*. OSF Preprints. <https://osf.io/preprints/edrxiv/93w6j>
- Bearman, M., & Ajjawi, R. (2024). When I say... artificial intelligence. *Medical education*, 58(11), 1273–1275. <https://doi.org/10.1111/medu.15408>
- Bearman, M., Fawns, T., Corbin, T., Henderson, M., Liang, Y., Oberg, G., ... Matthews, K. E. (2025). Time, emotions and moral judgements: how university students position GenAI within their study. *Higher Education Research & Development*, 1–15. <https://doi.org/10.1080/07294360.2025.2580616>
- Boud, D. (1995). Assessment and learning: Contradictory or complementary? In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 35–48). Routledge. <https://doi.org/10.4324/9780203062074-8>
- Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413. <https://doi.org/10.1080/02602938.2015.1018133>
- Broadfoot, P. M. (1996). *Education, assessment and society*. Open University Press.
- Bovill, C., Matthews, K. E., & Hinchcliffe, T. (2021). *Student partnerships in assessment* [Guidebook]. Advance HE. [https://s3.eu-west-2.amazonaws.com/assets.creode.advance-he-document-manager/documents/advance-he/AdvHE\\_Student%20Partnerships%20in%20Assessment\\_1633509850.pdf](https://s3.eu-west-2.amazonaws.com/assets.creode.advance-he-document-manager/documents/advance-he/AdvHE_Student%20Partnerships%20in%20Assessment_1633509850.pdf)
- Cheng, Y., Lyons, K., Chen, G., Gašević, D., & Swiecki, Z. (2024). Evidence-centered assessment for writing with generative AI. In *Proceedings of the 14th learning analytics and knowledge conference* (pp. 178–188)
- Cook-Sather, A. (2018). Tracing the evolution of student voice in educational research. In *Radical collegiality through student voice: Educational experience, policy and practice* (pp. 17–38). Springer Singapore.
- Cook-Sather, A. (2022). *Co-creating equitable teaching and learning: Structuring student voice into higher education*. Harvard Education Press.
- Cook-Sather, A., & Matthews, K. E. (2023). Practising student voice in university teaching and learning: Three anchoring principles. *Journal of University Teaching and Learning Practice*, 20(6), 1–11.
- Corbin, T., Dawson, P., Nicola-Richmond, K., & Partridge, H. (2025). ‘Where’s the line? It’s an absurd line’: towards a framework for acceptable uses of AI in assessment. *Assessment & Evaluation in Higher Education*, 50(5), 705–717. <https://doi.org/10.1080/02602938.2025.2456207>
- Corbin, T., Bearman, M., Boud, D., & Dawson, P. (2025b). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, 1–17. <https://doi.org/10.1080/02602938.2025.2553340>

- Corbin, T., Walton, J., Bannister, P., & Deranty, J. P. (2025c). On the essay in a time of GenAI. *Educational Philosophy and Theory*, 1–13. <https://doi.org/10.1080/00131857.2025.2572802>
- Farazouli, A., Cerrato Pargman, T., Bolander Laksov, K., & McGrath, C. (2025). Navigating uncertainty: university teachers' experiences and perceptions of generative artificial intelligence in teaching and learning. *Studies in Higher Education*, 1–16. <https://doi.org/10.1080/03075079.2025.2550766>
- 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>
- Lodge, J. M., Bearman, M., Dawson, P., Gniel, H., Harper, R., Liu, D., McLean, J., Ucnik, L., & Associates. (2025). *Enacting assessment reform in a time of artificial intelligence*. Tertiary Education Quality and Standards Agency, Australian Government.
- Lodge, J. M., Yang, S., Furze, L., & Dawson, P. (2023). It's not like a calculator, so what is the relationship between learners and generative artificial intelligence? *Learning: Research and Practice*, 9(2), 117–124. <https://doi.org/10.1080/23735082.2023.2261106>
- Luo, J., & Dawson, P. (2025). Exploring value judgements in grading: will teachers mark down student work assisted by GenAI, and should they? *Studies in Higher Education*, 1-15. <https://doi.org/10.1080/03075079.2025.2552825>
- Matthews K. E. (2018). Engaging students as participants and partners: An argument for partnership with students in higher education research on student success. *International Journal of Chinese Education*, 7(1), 42-64. <https://doi.org/10.1163/22125868-12340089>
- McArthur, J. (2018). *Assessment for Social Justice: Perspectives and Practices within Higher Education*. Bloomsbury Academic.
- Nicola-Richmond, K., Dawson, P., Partridge, H., & Macfarlane, S. (2025). It takes a village...Program-wide approaches to redesigning assessment in a time of generative artificial intelligence. *Journal of University Teaching and Learning Practice*, 22(7), <https://doi.org/10.53761/zpp2ja61>
- Purvis, A. J., Nicholas, V., & Tai, J. (Eds.). (2024). What's your problem?: Writing effective research questions for quality publications. *Journal of University Teaching and Learning Practice*, 21(10), 1-16.
- Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34(7), 807–826. <https://doi.org/10.1080/03075070802706553>
- Sadler, D. R. (2013). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5–19. <https://doi.org/10.1080/0969594X.2012.714742>
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., ... & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075.
- Walton, J., Bearman, M., Crawford, N., Tai, J., & Boud, D. (2025). How university students work on assessment tasks with generative artificial intelligence: matters of judgement. *Assessment & Evaluation in Higher Education*, 1–17. <https://doi.org/10.1080/02602938.2025.2570328>