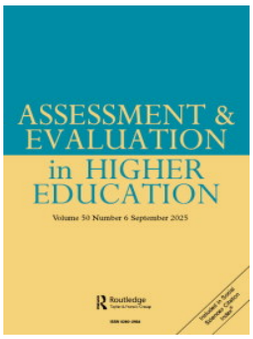

Assessment & Evaluation in Higher Education

ISSN: 0260-2938 (Print) 1469-297X (Online) Journal homepage: www.tandfonline.com/journals/caeh20

The wicked problem of AI and assessment

Thomas Corbin, Margaret Bearman, David Boud & Phillip Dawson

To cite this article: Thomas Corbin, Margaret Bearman, David Boud & Phillip Dawson (03 Sep 2025): The wicked problem of AI and assessment, Assessment & Evaluation in Higher Education, DOI: 10.1080/02602938.2025.2553340

To link to this article: https://doi.org/10.1080/02602938.2025.2553340

Published online: 03 Sep 2025.

ASSESSMENT & EVALUATION IN HIGHER EDUCATION
https://doi.org/10.1080/02602938.2025.2553340

Routledge
Taylor & Francis Group

The wicked problem of AI and assessment

Thomas Corbin, Margaret Bearman, David Boud and Phillip Dawson

Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Geelong, Victoria, Australia

ABSTRACT

Generative artificial intelligence (GenAI) has created significant assessment challenges in higher education. Universities and teachers have responded to these in various ways, including by way of new policies, revised assessment formats, and technological safeguards. These responses are typically predicated on an assumption that AI in assessment is a problem that can be solved if only the right approach can be found. However, in this paper we argue that GenAI may not be that kind of problem at all. Drawing on interviews with 20 teachers responsible for assessment design within a large Australian University, this paper applies Rittel and Webber framework of wicked problems to analyse educators' experiences with assessment design in the context of GenAI. Our findings demonstrate that the GenAI-assessment challenge exhibits all ten characteristics of wicked problems. For instance, it resists definitive formulation, offers only better or worse rather than correct solutions, cannot be tested without consequence, and places significant responsibility on decision-makers. In the light of this redefinition of the AI and Assessment problem, we argue that educators require certain institutional permissions – including permission to compromise, diverge, and iterate – to appropriately navigate the assessment challenges they face.

KEYWORDS

Generative AI; GenAI; assessment; educational technology

Introduction

Generative artificial intelligence (GenAI) challenges established assessment practices in higher education (Luo 2024; Jensen et al. 2024). Educators are grappling with how to preserve assessment validity and students appear concerned about being falsely accused of inappropriate AI use (Wu et al. 2024). Simultaneously, the new technology appears to offer powerful aids, potentially including improving assessment feedback or perhaps assisting students to better understand assessment expectations (Corbin, Tai, and Flenady 2025a; Sporrong, McGrath, and Cerratto Pargman 2025). The challenges for teachers and institutions exist not simply in policing misuse, but

CONTACT Thomas Corbin t.corbin@deakin.edu.au Centre for Research in Assessment and Digital Learning (CRADLE), Deakin University, Geelong, VIC, Australia.

in navigating the tensions between supporting productive, ethical use and maintaining valid, meaningful assessment (Fawns et al. 2024; Lodge et al. 2023).

There have been many responses to these challenges. Some focus on developing GenAI-oriented assessment policy (McDonald et al. 2025), others on detection tools (Ardito 2025). Some have emphasised the importance of clarity regarding AI use in assessment (Perkins et al. 2024), while others have advocated for restructuring assessment tasks (Corbin, Dawson, and Liu 2025b). These and similar approaches appear to be implicitly premised on the view that there is a 'correct solution' to the challenge of AI in assessment, a solution that we merely need to identify and then enact. However, if exploding levels of publication are any indication, the GenAI-assessment problem is not so easily solved. Instead, it may be the case that the challenge posed to assessment by GenAI is not that kind of problem at all.

In this paper we argue that there is reason to think the GenAI-assessment challenge exhibits the hallmarks of what Rittel and Webber (1973) term a 'wicked problem'. Wicked problems, as opposed to 'tame' problems, do not have 'correct' or 'incorrect' solutions (Rittel and Webber 1973). This does not mean there are no ways forward, nor does it mean that all ways forward are equally valuable. However, it does mean that responses must look very different. For one, they require a shift from seeking definitive answers to engaging in ongoing, adaptive work shaped by competing priorities and evolving conditions. Perhaps more importantly, it also means that the heavy burdens placed on those positioned to navigate the challenges, most obviously university teachers, are both substantial and – at least to a certain extent – unavoidable.

In what follows we draw on a qualitative interview study of university teachers that brings together the idea of wicked problems and teacher experience. To begin, in the following section, we provide an overview of wicked problems as a conceptual framework, outlining their defining characteristics and relevance to higher education. We then employ this as a deductive analytic frame to interrogate our application of wicked problem criteria as an analytical lens. The findings are then presented, organized around Rittel and Webber (1973) ten characteristics of wicked problems, illustrating how the reflections of teachers in charge of assessments (unit chairs) affirm the wicked nature of changing university assessments in response to AI. In the discussion, we explore the implications of these findings, arguing that understanding assessment as a wicked problem does not lead to paralysis, but instead offers a meaningful path forward.

On wicked problems

Wicked problems, as originally conceptualised by Rittel and Webber (1973), describe challenges that defy simple solutions. For problems of this kind there is no 'silver bullet', as Keep and Mayhew (2014, 765) put it, that will solve the problem. They are complex, contested, and continually evolving. Unlike their counterpart 'tame' problems, which have clear definitions and measurable solutions, wicked problems lack definitive formulations, and their solutions are not true or false but rather better or worse, requiring judgment, compromise, and adaptation. This distinction is key because it disrupts the assumption that there is a 'correct' policy, assessment method, or institutional response waiting to be discovered. Instead, every approach

carries trade-offs, is shaped by context, and must be continually reassessed in response to evolving conditions. For those tasked with navigating wicked problems, this reality has a significant personal toll; every decision feels provisional, every choice open to criticism, and the pressure to find the 'right' solution persists even when no such solution exists (Wexler 2009, 534).

Other complex and persistent challenges in higher education have been framed as wicked problems, including student engagement (Beaulieu and Roberge 2022) and student attrition (Kirk 2022). Some scholars have suggested that assessment design itself may exhibit the characteristics of a wicked problem (Canning and Eve 2020; Deeley et al. 2019), while Bearman et al. (2017) study of how university teachers design assessment suggest that this is a complex contextualised practice. Although they do not adopt a wicked problems framework, they do note that teachers, when designing assessments, bring their personal histories and disciplinary backgrounds to a particular local context. However, a challenge with the rise of GenAI is that one of teachers' traditional repertoire of responses based on their own past experiences is no longer meaningful in the same way.

The wicked problems framework is valuable not because it offers solutions but because it provides a more penetrating and practical way to think about problems that cannot be definitively solved. For example, in public policy, issues such as climate change (Austin and Háji 2024), social inequality (Kim et al. 2025), and urban planning (Asadi and Arbab 2024) have been productively addressed when framed as wicked, precisely because they involve competing priorities, incomplete information, and contested values. The concept has been particularly productive within organizational and management contexts, where appropriate decision-making and institutional change are often described as wicked problems which require flexibility, negotiation, and iterative adjustment (Head 2022).

Many influential institutional and scholarly responses to the GenAI-assessment problem in the higher education literature assume that the challenges it poses are tame rather than wicked. Recall that a tame problem is one that can be clearly defined and solved with the right intervention, for example through technological interventions. The clearest case of such an intervention is likely AI detection technology. The problem of AI and assessment, from at least one standpoint, is that for any given piece of writing teachers cannot tell what a student wrote themselves and what they outsourced to AI. If that is the problem, AI detection software could be seen as the correct solution, as, at least in theory, it can see beyond what the human teacher can and point them to areas of the text written by GenAI. Although he is sceptical regarding the likely success of these tools, as Ardito (2025) writes, 'Teachers naturally seek methods to control and guide AI usage in the classroom, and a working system to enable AI usage detection would be the most immediate, intuitive, and easy-to-implement solution to ensure academic integrity'. Underlying this view is an assumption that there is a problem – solution alignment that just needs to be uncovered. In this study, we look to teachers' accounts of assessment design to see if this is indeed the case.

This study examines whether and how the challenge of changing university assessments in response to GenAI aligns with the characteristics of a wicked problem. This framing has, to the best of our knowledge, not been used thus far with respect to

GenAI. We suggest that such an analysis matters precisely because misdiagnosing the problem type leads to counterproductive responses. Treating wicked problems as tame typically generates frustration, policy churn, and blame rather than progress. If the AI challenge is tame, then conventional solutions (such as stronger policies, better detection tools, or redesigned assessments) should eventually provide definitive resolution. But if it is wicked, then these approaches, while potentially valuable as partial responses, may be insufficient when positioned as complete solutions, potentially creating unrealistic expectations and cycles of policy frustration rather than progress.

Method

This paper draws on an analysis of interview data describing how university educators report their experiences of designing assessment that accounted for the presence of GenAI. We recruited participants by asking for referrals from key stakeholders (such as faculty Associate Deans of Teaching and Learning or equivalent) of those who they considered to be doing interesting assessment design work following the emergence of GenAI, and then inviting these teachers to participate. We were looking for participants such as this as we considered it likely that they consider there to be a problem that needed to be addressed, and so were likely to hold relevant opinions. We interviewed 20 of those responsible for running a subject ('Unit Chairs') at a large Australian university, evenly distributed across all faculties, helping to provide a diverse representation of disciplinary perspectives and assessment approaches.

Data collection occurred during the second half of 2024. We conducted semi-structured individual interviews remotely using Zoom. Each interview lasted approximately one hour. We prompted participants to provide detailed, reflective accounts of their experiences and decision-making processes. Major interview topics included teachers' conceptions of what constitutes 'appropriate' assessment, perceived impacts of GenAI on teaching and learning, specific adaptation strategies implemented, and the broader rationale guiding their assessment design choices. All interviews were audio-recorded with participant consent and transcribed verbatim. To promote confidentiality, identifying details were removed from transcripts, and participants were assigned pseudonyms. Ethical approval for the study was obtained (ID: 2024-098) and all participants provided informed consent.

We read and re-read our interview data and discussed it collectively. The overwhelming impression was how teachers were wrestling with the situations they found themselves in, and there were no easy assessment design answers, irrespective of how experienced the teachers were. This then led to a decision to analyse our interview data deductively, using a theoretical thematic analysis (Braun and Clarke 2006) guided by Rittel and Webber (1973) framework for characterising wicked problems. Interview responses were systematically mapped to these characteristics. This process was conducted manually by the researchers.

We acknowledge that while utilising the framework of wicked problems as a lens provides structured insights, it may also restrict the scope of analysis by excluding variables outside the model's parameters. This as an inherent trade-off in model-based research, that the theoretical clarity gained through a focused approach comes at the cost of potentially overlooking factors that lie beyond the given framework.

Nevertheless, for the purposes of this study, the model's boundaries enable us to isolate and examine the specific mechanisms of interest with the precision necessary to advance understanding of assessment in the context of GenAI. In the following section, we present this analysis.

Results

In this section, we proceed through each of Rittel and Webber (1973) ten characteristics. We first detail what each 'wicked problem' characteristic entails before examining to what extent teachers' experiences align with this conceptualization. Our analysis reveals that teachers' experiences with AI and assessment align strongly with all ten characteristics of wicked problems, confirming that this challenge is indeed better understood as a wicked rather than tame problem. Overall, while participants approached assessment change from different perspectives, their responses consistently revealed tensions, trade-offs, and uncertainties that resist simple resolution. The results highlight how teachers engage with these tensions in practice, sometimes finding ways to work within them, at other times encountering limits that constrain their ability to respond effectively. By tracing these patterns, we aim to provide insight into both the persistent difficulties of assessment change in response to AI and the adaptive strategies teachers employ in navigating them.

Wicked problems do not have a definitive formulation

The first defining feature of wicked problems is that they cannot be clearly or conclusively defined. Unlike technical problems where stakeholders can in theory agree on what needs fixing, wicked problems mean different things to different people and these varying definitions pull solutions in contradictory directions. Without agreement on what the problem is, a singular, cohesive response becomes impossible.

This lack of formulation was apparent in the markedly different ways participants framed the problem of designing assessment to accommodate the rise of GenAI. For some, the challenge was fundamentally about workforce readiness: 'If they're using it in the workforce already […] we can't just say to students, you cannot use it' (T5). This framing positioned resisting GenAI as professionally irresponsible with the corollary that the solution lies in integration into teaching. Another participant framed the issue very differently, casting GenAI use as a form of educational fraud: 'It's cheap learning, because students end up finishing university knowing zero, having learned zero from day one to the end' (T20). This framing positioned a response of restriction and policing, directly contradicting the integration approach.

Some teachers explicitly acknowledged that the problem contained layers of problems. For example, one stated that: 'We talk about GenAI as a problem of cheating, but it's also a problem of engagement, and it's also a problem of workload' (T3). Relatedly, when addressing similar tensions, another stated that 'It's hard, basically. I don't know. I'm at a loss. I keep on trying to have conversations with people, and people seem to be at a loss too' (T12). Other teachers also used this expression, finding themselves 'at a loss'. Consider for example the frustration of the teacher who stated: 'I've spent so much fucking time on developing this stuff. They're really good as units, things that

I'm proud of. Now I'm looking at what AI can do, and I'm like, what the fuck do I do? I'm really at a loss, to be honest'. (T10). Seen through a wicked problems lens, these teachers weren't failing to grasp the situation. Rather, they were confronting the difficulty of addressing a problem which resists singular definition.

They do not have a 'stopping rule' that signals when they are solved

The second defining characteristic of a wicked problem is that it has no stopping rule – that is, there are no clear criteria for knowing when you have reached 'the solution' (Rittel and Webber 1973 p. 160). Unlike solving a chess problem where checkmate signals definitive completion, or fixing a leaking pipe where success can be measured by the absence of water, wicked problems offer no such clarity. As Rittel and Webber present it, 'The planner terminates work on a wicked problem, not for reasons inherent in the "logic" of the problem. He stops for considerations that are external to the problem: he runs out of time, or money, or patience. He finally says, "That's good enough"'.

This absence of clear success criteria was evident throughout our data. Teachers consistently expressed uncertainty about whether their assessment adaptations had achieved their goals. When asked about determining success, one teacher responded: 'How do we actually tell? You can't' (T15). Another, who was similarly grappling with the possible changes they were encountering in assessment design, asked: 'Have I given enough warnings to students […] Have I ticked off all the boxes […] Have I struck that right balance? I don't know' (T6).

These responses illustrate how teachers could always refine their assessments further – adding more security measures, more authentic elements, more safeguards – but lack definitive criteria for when their design work is complete. They stop not because they have reached an optimal solution, but because of time constraints, institutional deadlines, or simply the need to move forward with something workable.

Their solutions are not true or false, only good or bad

Technical problems have correct answers that can be verified. Wicked problems, on the other hand, have only trade-offs, where every response sacrifices something valuable. Understanding this reframes 'failure' as inherent to the structure of the problem rather than inadequacy of the solution.

For participants, these trade-offs manifested practically in assessment design. One teacher created dual submission streams – one with and another without AI – in an attempt to balance competing learning objectives. According to the teacher, this approach aimed to preserve student agency by letting students choose whether or not to engage with AI, with each stream assessed using tailored rubrics to reflect different expected skill sets. The result: 'It also created a huge amount of work, additional work, because it was effectively two assessments and it was a bit of a nightmare' (T12). This dual track solution 'worked better' in serving pedagogical goals but 'worked poorly' by creating unsustainable workload. Another unit chair worried: 'We can make assessments more AI-proof, but if we make them too rigid, we just test compliance rather than creativity' (T3).

These types of statements illustrate how moves toward assessment security sacrifice something else, be it authenticity, creativity, or real-world relevance. If there is a correct answer, then teachers are simply failing in their duty to find the right balance. But, from a wicked problems perspective, they are navigating a landscape where potentially every choice is simultaneously both right and wrong.

There is no way to test the solution to a wicked problem

Rittel and Webber (1973) argue that wicked problems lack clear metrics for testing whether solutions have succeeded. Our data revealed multiple ways teachers experienced this absence of verifiable outcomes in their assessment practices.

Several unit chairs expressed uncertainty about whether their assessment adaptations were effective. When asked about determining success, one stated simply: 'If a student uses AI appropriately for brainstorming, we might never know. If they use it inappropriately, we also might never know' (T18). Attempts to gain certainty through technological means proved similarly frustrating. One unit chair recounted testing AI detection software with their own writing: 'There's no point putting anything through AI to say, 'Did you write this?' Because we tested the system already and we wrote something and we put it through AI and said, 'Did you write this?' They said "Yes". We go, "Well, no. You didn't because we did". You can't even trust that sort of thing' (T4).

This inability to verify whether approaches were working or not appeared across different assessment strategies. Teachers could not determine if their adaptations successfully preserved academic integrity or enhanced learning or prepared students appropriately.

They cannot be studied through trial and error, 'every trial counts'

The fifth characteristic of wicked problems is that solutions cannot be found through experimenting with solutions because every attempt has real consequences. We identified this characteristic in our data where teachers described the high stakes attached to their assessment decisions.

Several participants, for example, expressed concern about the lasting impacts of their choices. These typically fell into consequences falling on themselves as teachers, on students, or on the university itself. For example, whilst discussing the aim of coming up with novel, AI appropriate assessments, one teacher voiced their concern about how it might impact student enrolments: 'I also think that if I'm the only one [changing essays for more novel assessments], I would be punished. I will have less and less students enrolling in my units. Potentially, my units will die out because they would be more difficult. Of course, I have a career, I have a family and a mortgage. So, I don't want to do it by myself.' (T20) Teachers also focused on institutional reputation. One teacher, for instance, told us: 'That is my biggest concern because at the end of the day, it's the reputation of the university, not necessarily the student. It could affect the [University] name because we're sending students out into the workforce who potentially haven't completed a lot of their degree' (T3). The irreversible nature of educational experiences was a recurring theme. Some

teachers framed AI as an existential challenge to education itself: 'I'm also thinking, at the same time, how long have we got? How long have we got while we're still institutions that students come into that we are in the position to do this before AI might do it for us?' (T12). The same participant asked: 'How long are we a legitimate institution? How long do we remain justifiable?'

These accounts illustrate how assessment design decisions in the AI era cannot be treated as low-stakes experiments. Teachers cannot simply trial different approaches to see what works because each attempt carries potential consequences for their careers, their students' learning, and their institutions' reputations.

There is no end to the number of solutions or approaches to a wicked problem

Rittel and Webber (1973) next propose that wicked problems present limitless possible approaches with no way to determine if all options have been considered. This vast array of potential responses is reflected in the breadth of strategies teachers described in response to AI. Some of these were proposed, and some were already implemented.

Participants described numerous different approaches to assessment adaptation. The range of specific strategies mentioned included oral examinations, portfolio assessments, industry partnerships, peer review systems, in-class presentations, authentic tasks, and various hybrid approaches. One teacher expressed preference for traditional methods: 'I know I would still prefer exams to come back on campus because it would be the only piece of assessment that we can truly say this is their own work' (T4). Some advocated for embracing AI as a professional tool: 'I think GenAI is going to stay, right? It's already part of the workforce, like us as well. Students need to be able to use it efficiently' (T17). Others explored increased control measures: 'I think we're probably going to have to involve some form of invigilation… I personally think we do need to figure out some way of controlling it' (T5). Still others sought middle ground through selective integration: 'The assessment for them to actually be useful and authentic, we need to at least somewhere in the course actually incorporate GenAI in assessment' (T6). This variety existed both across interviews and sometimes within single conversations, with teachers describing multiple potential approaches without clear criteria for selection among them.

All wicked problems are essentially unique

The appeal of standardized solutions - whether "best practice" templates or institutional mandates - assumes that similar-looking problems can be solved with similar approaches. But wicked problems resist this logic because each instance emerges from an irreducibly specific context. This characteristic appeared in our data through teachers' descriptions of how approaches that worked in one context failed in others.

The context-dependent nature of solutions was particularly evident for oral assessments. One philosophy teacher described positive experiences: 'It was really nice, actually, to see in the two classes, students, doing philosophy, talking to each other

about concepts about the limitation of particular arguments or ideas' (T1). However, another teacher calculated the practical implications of such an assessment for the larger cohorts typical within their university: '250 students by five minutes, make it 10 min […] that's, yes, it's like 2500 min, and then that's how many days of work is it just to administer one assessment?' (T12).

Institutional constraints created additional variation. One teacher reported attempting to implement invigilation but encountering resistance: 'The university fought us on it. They wouldn't even give us the rooms to have exams invigilated. … or the computers, or anything' (T9). Disciplinary differences also emerged as significant. While some teachers in business-oriented fields emphasized AI preparation, with one teacher stating that 'students need to be able to use it efficiently' (T17), others in clinical fields raised different concerns. A nursing teacher considered dealing with the assessment challenge by simply removing all written assessments, asking 'Do nurses need to know how to write?' (T14), indicating how professional contexts also shaped assessment priorities.

Participants frequently noted how their specific combination of factors, whether that be discipline, cohort size, institutional resources, professional requirements, each created distinct challenges. Responses mentioned by colleagues in other faculties or disciplines often proved impractical or inappropriate when teachers considered their own contexts.

Wicked problems can always be described as the symptom of other problems

Wicked problems do not exist in isolation but instead emerge from and reveal deeper structural issues. Rittel and Webber (1973) note that wicked problems are interconnected with other systemic issues. In our interviews, teachers consistently described AI challenges as symptoms of deeper structural problems in higher education.

Several participants saw AI vulnerabilities as symptoms of institutional business models. One teacher argued: 'A university like [the one in which I work], which is based on a business model, which is online-based, where you cannot incentivize students to come in person, and all the assessments are based on tasks you ask students to do at home in their own time, this model is the most vulnerable to fraud in an age of AI' (T9). They concluded that this model led to 'cheap learning, because students end up finishing university knowing zero, having learned zero from day one to the end'. Others described AI as a symptom of engagement crises. One teacher linked this concern to a related problem of engagement: 'the consumer-driven model of education learning […] basically provides a justification for non-engagement and that then with GenAI coming into this' (T1). This manifested in transactional approaches where 'The vast majority are just worried about the assessment. They want to pass the unit' (T3). This concern was frequently raised, for example by one teacher who stated: 'There's not as many students that are actually there to learn, they're just there to pass their exam. That's why cheating [with AI] is a big problem' (T11).

AI challenges also appeared as symptoms of workload pressures. One teacher stated that: 'If [my University] wants us to do this in a really effective and purposeful way, the first step would be to give people time to do that assessment planning to revise their units' (T16). Another noted similar resource constraints:

'We just do not have the resources to be able to detect everything and then to write up any breaches' (T12). Relatedly, many teachers commented on the problem of AI and assessment being one of time, with many teachers repeatedly described feeling unable to keep pace with AI's evolution. One stated directly: 'I'm running to stand still' (T18), while others expressed similar frustrations: 'AI is moving too quickly for universities. We can't keep up' (T11) and 'AI moves so much quicker than we're moving' (T3). Another captured this ongoing adaptation challenge: 'Every time I think I've adjusted the assessments to make them AI-resistant, AI improves' (T15).

Across these accounts, the problem of AI in assessment frequently emerged as a symptom of pre-existing conditions: business models or insufficient time and resources for teaching or educational cultures that had already become transactional before AI arrived.

The way a wicked problem is described determines its possible solutions

The ninth characteristic of a wicked problem is that the way the problem is framed shapes which solutions become possible. This relies on the claim that how we define a problem constrains what kinds of responses can be imagined or pursued. In other words, how we frame the AI and assessment challenge predetermines which solutions appear reasonable and which remain invisible.

Our data demonstrated how different framings led to radically different, often contradictory, responses. When teachers framed AI as a threat to academic integrity, they favoured control-based solutions. One stated: 'I know I would still prefer exams to come back on campus because it would be the only piece of assessment that we can truly say this is their own work' (T4). Another who saw AI as enabling cheating advocated for 'some form of invigilation' (T5). Those who framed AI as a professional necessity proposed integration: 'I think GenAI is going to stay, right? It's already part of the workforce, like us as well. Students need to be able to use it efficiently. The part of their skills they will need to learn would be to use GenAI efficiently' (T17). When participants framed the issue as a workload problem, they focused on addressing this issue alone: 'If [the University] wants us to do this in a really effective and purposeful way, the first step would be to give people time to do that assessment planning to revise their units' (T16).

Different framings appeared to lead toward different solution types: control measures for integrity threats, integration for professional tools, resources for workload issues, and fundamental questioning for existential concerns.

Those who present solutions to these problems have no right to be wrong

The tenth characteristic of wicked problems is that decision-makers bear full responsibility for the consequences of their choices. Unlike theoretical problems where errors have no real-world impact, those addressing wicked problems are, as Rittel and Webber (1973, 167) note, 'liable for the consequences of the solutions they generate'.

This liability manifested in multiple ways across our interviews. One teacher worried about graduating unprepared professionals: 'How many are we missing?

Are we in fact sending students out into the workforce who can get through an interview, but when they start doing the job, they can't?' (T11). The personal vulnerability this created was articulated starkly: 'I feel very, very vulnerable within the university running assessments like this because I know that there are pockets of the university management who would really like to just see us do traditional, detached, academic assessments that don't threaten to push students' (T6).

This characteristic of bearing responsibility for consequences while lacking control over outcomes appeared to create significant stress for teachers navigating assessment decisions in the context of AI.

Discussion

Our findings show that changing university assessments in response to AI exhibits all ten characteristics of a wicked problem. This recognition suggests that we need to reframe the challenge. In the context of higher education assessment, the challenge posed by GenAI is better understood not as a problem to be solved but as a condition to be navigated.

Teachers in this study described first-hand how designing assessment has no simple solution. They wrestled with the challenge in front of them. Their experiences suggest that no assessment model or policy eliminates the fundamental tensions between academic integrity, student learning, workload, and evolving technology. Instead, assessment in the age of GenAI is an ongoing negotiation, requiring continual adaptation rather than prescriptive, one-size-fits-all solutions. But, this does not lift the burdens associated with the problem. On the contrary, it places them squarely on the shoulders of teachers who must navigate conflicting guidance, constrained resources, and competing expectations while trying to design assessments that are valid assurances of learning (Corbin et al. 2025c; Dawson et al. 2024) and at the same time also appropriately prepare students to live in a world where AI is ubiquitous (Lodge et al. 2023).

Using the frame of wicked problems highlights useful responses and areas where teachers are struggling. Teachers' strategies that addressed the wicked problem characteristics directly included: Acknowledging the multitudes of problems and responses; understanding that constant revision was necessary; negotiating between various outcomes; and acknowledging the local nature of their particular problem. Where they struggled was around the real challenge of managing uncertainty and bearing responsibility for potentially deleterious consequences. Some were overwhelmed by options. There was a sense, that many, similarly to Sporrong, McGrath, and Cerratto Pargman (2025) participants, felt that better solutions were possible, but that such solutions could not be achieved given the problems of time or low resource or a transactional education system.

This is where the value of an explicit wicked problem frame may assist. If we frame AI as a tame problem, then educators are simply failing by implementing the wrong solutions until the right one is found. But if it can be recognised as a wicked problem, as our data suggests, then educators are engaged in something

fundamentally different. They are put in the position of needing to make continuous professional judgments in conditions of permanent uncertainty. This imposes a burden that is both substantial and yet at the same time necessary. This kind of burden demands acknowledgement and appropriate consideration.

Designing university assessment was a complex, situated affair, 'working in complex, often overwhelming, environments' (Bearman et al. 2016, 553), even before the rise of GenAI. Teachers treat, for the most part, assessment as a professional practice (Norton, Floyd, and Norton 2019); this implies that they must make professional judgements to manage the multiple purposes of assessment that are often in tension (Boud & Associates 2010) The GenAI challenge intensifies this existing responsibility rather than creating something wholly new. The unit chair asking, 'Do nurses need to know how to write?' is not facing a novel burden imposed by GenAI. They are exercising the same professional judgment about assessment teachers have always needed to answer: What skills or knowledge matter most for these students, in this profession, at this moment? GenAI makes the question more urgent and complex – and as argued, the repertoire of design responses has been challenged – but the work of assessment design itself remains as it always does?

So, what can be created if we reframe the GenAI-assessment challenge as a wicked problem? First, it lifts the impossible burden on teachers and institutions to immediately get things right once and for all. When problems are unsolvable and ever-changing, missteps and course corrections are not failures. They are part of doing the work well. More importantly, a wicked problem framing reorients assessment design itself. It shifts the role of educators from implementing fixed solutions to engaging in a continual search for better, context-sensitive designs that respond to local needs, disciplinary values, and the evolving presence of AI in student learning. This shift does not merely legitimise experimentation, though that is one outcome. It opens up the possibility of more meaningful, inclusive, and future-relevant assessment practices. It makes space for assessments that prioritise assuring learning over policing, that are resilient in the face of technological change, and that reflect the diversity of student experience. It allows institutions to value professional judgment, to tolerate uncertainty, and to build systems that evolve rather than lock in a solution. In short, reframing the problem does not solve it, but it creates the conditions under which both problem navigation *and* better assessment becomes possible.

How do we get from here to there? To begin, we turn back to a study of assessment design that identified that those who negotiated the challenge of assessment design easily, did so using what was labelled a meta-design (Bearman et al. 2017) process. That is, these teachers consciously articulated 'what was mutable or negotiable within their contexts generally and what compromises could and couldn't be made surrounding their assessment design and implementation'. We turn here to three meta-design strategies. These are 'permissions' that teachers can give themselves and which institutions can seek to support. Each permission helps surmount the ideas of solutionism that our data suggests prevents educators from exercising the professional judgment they cannot avoid and should not surrender.

Permission to compromise

If we think of AI and assessment as a technical problem, it is easy for us to fall into the trap that we must find 'the solution' in our classroom which perfectly demonstrates success across all dimensions simultaneously; assessments must be secure *and* authentic *and* efficient *and* transformative. Alternatively, a wicked problem frame suggests that trade-off are necessary and there is no optimal balance nor solution. The teacher who wondered 'Have I struck the right balance? I don't know' (T6) was describing the uncertainty inherent in weighing pedagogical goals against workload, security against authenticity, current needs against future preparation. Permission to compromise recognises that wicked problems force impossible choices. Every assessment design (with or without GenAI) involves trade-offs between equally important values: security versus authenticity, workforce preparation versus foundational learning, manageability versus richness.

By granting ourselves permission to compromise, we acknowledge two realities. First, that every approach sacrifices something valuable. The assessment that best prevents AI exploitation might sacrifice creativity as much as the authentic task that engages students might quadruple marking time, whilst the contextually perfect solution might prove impossible to scale across the institution. Second, that some approaches will simply fail, generating valuable knowledge about what doesn't work in specific contexts. This permission removes the toxic burden of pursuing perfection that cannot exist. It allows educators to state plainly that this assessment prioritizes X at the expense of Y, and here is why. It transforms institutional culture from one that punishes imperfection to one that learns from it. When we stop seeking perfect solutions, we can start having honest conversations about which trade-offs serve our students best, which failures taught us most, and how to be thoughtfully imperfect rather than accidentally inadequate.

Permission to diverge

Permission to diverge recognizes that in wicked problems, context determines everything. What transforms learning in a 20-student philosophy seminar becomes logistically impossible with 250 business students. What prepares future lawyers for AI-integrated practice might undermine the clinical skills nurses need.

At its core, 'permission to diverge' means accepting that successful practices in one educational context need not – and often should not – be replicated elsewhere. It is the recognition that divergent approaches to common challenges can reflect contextual wisdom rather than inconsistency or failure. By granting ourselves permission to diverge, we acknowledge that different contexts might require quite different responses. This recognises that quality manifests differently across years, disciplines, cohort sizes, and professional destinations. The business educator who integrates AI because employers demand it and the nursing educator who restricts it to ensure clinical competence are both appropriate. Divergence can reflect wisdom that we can easily mistake for confusion. This permission transforms institutional expectations from uniformity to fitness for purpose. Divergence becomes a sign of thoughtful response rather than institutional failure. When faced with wicked problems, where context determines viability, this

permission to diverge releases us from the expectation that good solutions must be universally applicable.

Permission to iterate

Assessment design by its nature is iterative, not a one-time creation (Bearman et al. 2017). But these are heightened times. When AI capabilities transform monthly, when student behaviours shift each semester, and when professional requirements evolve constantly, the result can be that educators design assessments for yesterday's technology, implemented with today's students, preparing for tomorrow's unknowns. Permission to iterate recognizes that wicked problems evolve continuously, making fixed solutions obsolete.

The permission to iterate recognizes that wicked problems evolve continuously, making fixed solutions obsolete. From a teacher perspective it could mean building revision into workload models rather than treating it as failure of initial design or an unusual event. It likely means creating systems that support rapid adaptation rather than punishing frequent change. It certainly means expecting assessments to evolve rather than demanding they remain stable. When educators can adjust based on what they learn (that dual submission created impossible workload, that AI capabilities exceeded expectations, that students completed tasks in unexpected ways) they respond to reality rather than defending outdated plans. This permission transforms assessment from a product to be delivered to a practice to be refined (Boud et al. 2018).

These three permissions are particularly heightened in face of the GenAI-Assessment problem but, as others have noted, assessment design itself is essentially a wicked problem (Canning and Eve 2020). The design repertoire of assessment practices in the new GenAI era may well stabilise into formats or tasks that can be adapted to local contexts - what Goodyear (2005) might call pattern designs. But the key to these pattern designs is that they are not solutions, but rather provide a shorthand way to assist educators in their professional practice. And we propose that these permissions – to compromise, to diverge and to iterate – are useful meta-design strategies that will continue to help teachers adapt to the next great assessment crisis.

Conclusion

Universities that continue to chase the elusive 'right answer' to AI in assessment will exhaust their educators while failing their students. Those that embrace the wicked nature of this problem can build cultures that support thoughtful professional judgment rather than punish imperfect solutions.

The path forward requires abandoning the search for silver bullets in favour of developing adaptive capacity. This means creating institutional structures that support educator decision-making rather than mandating uniform responses, recognizing divergent approaches as evidence of contextual wisdom rather than institutional inconsistency, and treating assessment iteration as professional development rather than design failure. Of course, acknowledging the kind of uncertainty that comes with a wicked problem is difficult, and might even seem paralysing, tantamount to saying 'it's all too complicated'. Indeed, many of our participants appear paralysed,

right now, overwhelmed with the burden of what lies in front of them. However, while it is true that wicked problems do not have correct solutions, they do have better and worse responses. Removing the spectre of 'finding the perfect solution' just might help teachers navigate AI related challenges in more sustainable, healthy, and effective ways.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Thomas Corbin is a Research Fellow in the Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University. His research focuses on GenAI, assessment, and the philosophy of education.

Margaret Bearman is Professor in the Centre for Research in Assessment and Digital Learning, Deakin University. She researches assessment, feedback and digital learning in higher and professional education and is a Higher Education Academy Senior Fellow.

David Boud is Alfred Deakin Professor and Foundation Director of the Centre for Research in Assessment and Digital Learning at Deakin University, Australia he is also Emeritus Professor at the University of Technology Sydney and Professor of Work and Learning at Middlesex University. He is an Australian Learning and Teaching Senior Fellow.

Phillip Dawson is a professor and Co-Director of the Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University. His research focuses on assessment, feedback, and cheating.

ORCID

Thomas Corbin http://orcid.org/0000-0001-8750-5711
Margaret Bearman http://orcid.org/0000-0002-6862-9871
David Boud http://orcid.org/0000-0002-6883-2722
Phillip Dawson http://orcid.org/0000-0002-4513-8287

References

Ardito, C. G. 2025. "Generative AI Detection in Higher Education Assessments." *New Directions for Teaching and Learning* 2025 (182): 11–28. https://doi.org/10.1002/tl.20624.

Asadi, D., and P. Arbab. 2024. "Developing a Framework to Identify Wicked Problems in Urban Planning." *Geographical Urban Planning Research (GUPR)* 12 (1): 21–41.

Austin, Z., and A. Háji. 2024. "What Could (Or Should) Be the Regulatory Response to the Wicked Problem of Climate Change?" *Journal of Medical Regulation* 110 (2): 7–13. https://doi.org/10.30770/2572-1852-110.2.7.

Bearman, M., P. Dawson, S. Bennett, M. Hall, E. Molloy, D. Boud, and G. Joughin. 2017. "How University Teachers Design Assessments: A Cross-Disciplinary Study." *Higher Education* 74 (1): 49–64. https://doi.org/10.1007/s10734-016-0027-7.

Bearman, M., P. Dawson, D. Boud, S. Bennett, M. Hall, and E. Molloy. 2016. "Support for Assessment Practice: Developing the Assessment Design Decisions Framework." *Teaching in Higher Education* 21 (5): 545–556. https://doi.org/10.1080/13562517.2016.1160217.

Beaulieu, D., and J. Roberge. 2022. "The Wicked Problem of Engaging Students in Their Studies." *Pédagogie Collégiale* 35 (3): 7–14.
Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.
Boud, D, and Associates. 2010. *Assessment 2020: Seven Propositions for Assessment Reform in Higher Education*. Sydney: Australian Learning and Teaching Council.
Boud, D., P. Dawson, M. Bearman, S. Bennett, G. Joughin, and E. Molloy. 2018. "Reframing Assessment Research: Through a Practice Perspective." *Studies in Higher Education* 43 (7): 1107–1118. https://doi.org/10.1080/03075079.2016.1202913.
Canning, J., and J. Eve. 2020. "Assessment in Higher Education: The Anatomy of a Wicked Problem." In *Research Matters: The Pedagogic Research Conference and Articles 2019* (pp. 51–60). Brighton: University of Brighton.
Corbin, T., Tai, J., & Flenady, G. (2025a). Understanding the place and value of GenAI feedback: a recognition-based framework. *Assessment & Evaluation in Higher Education*, *50*(5), 718–731. https://doi.org/10.1080/02602938.2025.2459641
Corbin, T., P. Dawson, and D. Liu. 2025b. "Talk is Cheap: Why Structural Assessment Changes Are Needed for a Time of GenAI." *Assessment & Evaluation in Higher Education:* 1–11. https://doi.org/10.1080/02602938.2025.2503964.
Corbin, T., P. Dawson, K. Nicola-Richmond, and H. Partridge. 2025c. "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.
Dawson, P., M. Bearman, M. Dollinger, and D. Boud. 2024. "Validity Matters More than Cheating." *Assessment & Evaluation in Higher Education* 49 (7): 1005–1016. https://doi.org/10.1080/02602938.2024.2386662.
Deeley, S. J., M. Fischbacher-Smith, D. Karadzhov, and E. Koristashevskaya. 2019. "Exploring the 'Wicked'problem of Student Dissatisfaction with Assessment and Feedback in Higher Education." *Higher Education Pedagogies* 4 (1): 385–405. https://doi.org/10.1080/23752696.2019.1644659.
Fawns, T., M. Henderson, K. Matthews, G. Oberg, Y. Liang, J. Walton, T. Corbin, M. Lynn Bearman, S. B. Shum, et al. 2024. "Gen AI and Student Perspectives of Use and Ambiguity: A Multi-Institutional Study." In *Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education 2024: Navigating the Terrain: Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies*
Goodyear, P. 2005. "Educational Design and Networked Learning: Patterns, Pattern Languages and Design Practice." *Australasian Journal of Educational Technology* 21 (1): 82–101. https://doi.org/10.14742/ajet.1344.
Head, B. W. 2022. *Wicked Problems in Public Policy: Understanding and Responding to Complex Challenges*, 176. Springer Nature.
Jensen, L. X., A. Buhl, A. Sharma, and M. Bearman. 2024. Generative AI and Higher Education: A Review of Claims from the First Months of ChatGPT. *Higher Education* 89: 1145–1161.
Keep, E., and K. Mayhew. 2014. "Inequality–'Wicked Problems', Labour Market Outcomes and the Search for Silver Bullets." *Oxford Review of Education* 40 (6): 764–781. https://doi.org/10.1080/03054985.2014.979580.
Kim, S., H. Kwak, E. Nwakpuda, A. Sullivan, and V. M. Fenley. 2025. "Social Equity for Wicked Problems: Achieving Racial Equity in Homeless Service Provision." *Public Management Review* 27 (2): 395–427. https://doi.org/10.1080/14719037.2023.2246969.
Kirk, G. 2022. "How Can we Address the Wicked Problem of University Student Attrition?: A BEd (Early Childhood) Case Study." *Issues in Educational Research* 32 (3): 982–1000.
Lodge, J. M., S. Howard, M. Bearman, and P. Dawson Associates. 2023. & *Assessment Reform for the Age of Artificial Intelligence*. Australia: Tertiary Education Quality and Standards Agency.

Luo (Jess), J. 2024. "A Critical Review of GenAI Policies in Higher Education Assessment: A Call to Reconsider the "Originality" of Students' Work." *Assessment & Evaluation in Higher Education* 49 (5): 651–664. https://doi.org/10.1080/02602938.2024.2309963.

McDonald, N., A. Johri, A. Ali, and A. H. Collier. 2025. "Generative Artificial Intelligence in Higher Education: Evidence from an Analysis of Institutional Policies and Guidelines." *Computers in Human Behavior: Artificial Humans* 3: 100121. https://doi.org/10.1016/j.chbah.2025.100121.

Norton, L., S. Floyd, and B. Norton. 2019. "Lecturers' Views of Assessment Design, Marking and Feedback in Higher Education: A Case for Professionalisation?" *Assessment & Evaluation in Higher Education* 44 (8): 1209–1221. https://doi.org/10.1080/02602938.2019.1592110.

Perkins, M., L. Furze, J. Roe, and J. MacVaugh. 2024. "The Artificial Intelligence Assessment Scale (AIAS): A Framework for Ethical Integration of Generative AI in Educational Assessment." *Journal of University Teaching and Learning Practice* 21 (06): 49–66. https://doi.org/10.53761/q3azde36.

Rittel, H. W., and M. M. Webber. 1973. "Dilemmas in a General Theory of Planning." *Policy Sciences* 4 (2): 155–169. https://doi.org/10.1007/BF01405730.

Sporrong, E., C. McGrath, and T. Cerratto Pargman. 2025. "Situating AI in Assessment—An Exploration of University Teachers' Valuing Practices." *AI and Ethics* 5 (3): 2381–2394. https://doi.org/10.1007/s43681-024-00558-8.

Wexler, M. N. 2009. "Exploring the Moral Dimension of Wicked Problems." *International Journal of Sociology and Social Policy* 29 (9/10): 531–542. https://doi.org/10.1108/01443330910986306.

Wu, C., X. Wang, J. Carroll, and S. Rajtmajer. 2024. Reacting to Generative AI: Insights from Student and Faculty Discussions on Reddit. In Proceedings of the 16th ACM Web Science Conference (pp. 103–113).