Assessment reform for the age of artificial intelligence

November 2023

TEQSA

Contents

Assessment reform for the age of artificial intelligence

Jason M Lodge, The University of Queensland
Sarah Howard, University of Wollongong
Margaret Bearman, Phillip Dawson, Deakin University

With **Shirley Agostinho**, University of Wollongong, **Simon Buckingham Shum**, University of Technology Sydney, **Chris Deneen**, University of South Australia, **Cath Ellis**, The University of Sydney, **Tim Fawns**, Monash University, **Helen Gniel**, TEQSA, **Rowena Harper**, Edith Cowan University, **Michael Henderson**, Monash University, **Danny Liu**, The University of Sydney, **Lina Markauskaite**, The University of Sydney, **Jan McLean**, University of Technology Sydney, **Carlo Perrotta**, The University of Melbourne, **Lambert Schuwirth**, Flinders University, **Christine Slade**, The University of Queensland

This document was commissioned by TEQSA to support institutions as they reflect on the impact of generative artificial intelligence on assessment practices. The advice contained within this document is not part of TEQSA's suite of Guidance Notes and is not intended to be prescriptive. There will likely be considerable variation in how individual institutions approach assessment reform implementation. Assessment is always complex and assessment design entails compromise. Thus, we offer inspiration for teaching and learning leadership, avoiding formulaic solutions but providing expert insights into how and why assessment may need to change in an age of artificial intelligence.

Preamble

The emergence of generative artificial intelligence (AI), while creating new possibilities for learning and teaching, has exacerbated existing assessment challenges within higher education. However, there is considerable expertise, based on evidence, theory and practice, about how to design assessment for a digital world, which includes artificial intelligence. AI is not new, after all, even if the current iterations of generative AI are. This document, constructed through expert collaboration, draws on this body of knowledge and outlines directions for the future of assessment. It seeks to provide guidance for the sector on ways assessment practices can take advantage of the opportunities, and manage the risks, of AI, specifically generative AI.

The most immediate concern is that students may use generative AI in assessment tasks, calling into question their personal learning attainment. It is challenging to design non-invigilated assessment tasks that preclude substantial use of generative AI, and it appears almost impossible to detect if these technologies have been used in the production of assessment products in a reliable way. Moreover, there is every possibility that what generative AI can produce will be of a passable quality for many assessment tasks, particularly as student capabilities for using these tools evolve over time. Therefore, it is necessary to address the nature of assessment in relation to generative AI directly to provide students and teachers with ways to approach this issue productively.

Collectively, these threats to academic integrity highlight weaknesses in traditional approaches to assessment. While some forms of assessment seem to be robust to the

threat posed by generative AI, many assessment tasks are at risk. There is little value in ignoring AI or implementing blanket bans on particular tools or technologies. These are oversimplified solutions to a complex set of problems and overlook what is already known about good assessment practice. As AI use becomes commonplace across schools and workplaces, it will be increasingly important to consider how these tools are integrated into learning and teaching in higher education in intelligent ways.

We take our starting point for this document from the propositions for assessment outlined in Assessment 2020 (Boud and Associates, 2010). That foundational work outlines how assessment acts as a powerful intervention in student learning and highlights the educational purposes of assessment in parallel with the process of assuring learning outcomes. Good assessment design that allows for 'rich portrayals' of student learning is critical. Thus, we take as given that assessment should engage students in learning, provide a partnership between teachers and students, and promote student participation in feedback. These key elements of assessment can then guide how best to consider the role of AI in assessment design.

In addition to the foundation of Assessment 2020, we also align specifically with the *Higher Education Standards Framework (Threshold Standards) (2021)* for assessment (particularly section 1.4) and academic integrity (section 5.2). Moreover, we adhere to core principles underpinning an inclusive higher education system (see also section 2.2). These foundational principles include fairness, accessibility, transparency, equity, a right to privacy, and respecting students as individuals among diverse student cohorts, in alignment with the school sector (as per National Artificial Intelligence Taskforce, 2023). The right of students to be partners in their learning underpins all representations in this document. Given this position, we highlight the importance of ensuring learners are not disadvantaged by the inclusion of AI in assessment.

The development of AI technologies is proceeding at an unprecedented rate. This is therefore a document that is intended to be collaboratively reviewed and renegotiated with a range of stakeholders, including professional accreditation bodies. Our collective aim is to commence charting general directions forward, based on expertise, rather than predict every possibility on the horizon or provide prescriptive guidance. This is not a *map* of what may lie ahead but rather a *compass* to guide appropriate directions for assessment practices in higher education.

Guiding principles

These principles capture the essence of the considerations that are required for higher education assessment and AI. These principles, in combination with the important fundamentals described in the preamble, form the foundation on which the key propositions are built.

1. Assessment and learning experiences equip students to participate ethically and actively in a society where AI is ubiquitous

AI represents an urgent catalyst for change. It does not just influence how student learning can be assessed, it also influences what is worth assessing and, consequently, what and how students learn. This necessarily includes the ability to use AI tools, as well as a broader understanding of the ethics, limitations, biases, and implications of AI. While acute issues relate to academic integrity, the longer-term challenge for assessment design will be to incorporate these new technologies into higher education in a thoughtful and evidence-informed manner.

Responding to the risk posed by generative AI needs to focus not only on what is inappropriate but also on what is appropriate.

2. Forming trustworthy judgements about student learning in a time of AI requires multiple, inclusive and contextualised approaches to assessment

There is no single assessment type that can account for all desirable and undesirable uses of AI by students. Using multiple assessments of different types, when triangulated, provides greater trustworthiness and allows for practices that are more inclusive. This principle captures an approach to assessment design that is increasingly necessitated by the introduction of generative AI.

Propositions

The following propositions elaborate on the high-level principles and outline aspirational emphases for higher education assessment in a world where students will engage with AI, including generative AI, during their studies and lives.

Assessment should emphasise...

1. ...appropriate, authentic engagement with AI

Assessment should encourage students to critically analyse AI's role in, and value for, work and study, aligned with disciplinary or professional values. Assessment tasks should be designed to foster responsible and ethical use of AI in ways that are authentic to both the task and the discipline. Such engagement should be meaningful, supported through explicit teaching across a program of study, and aligned with the program learning outcomes. What is appropriate will vary across time and place.

Learning to work with AI is rapidly becoming a critical capability for many, if not most, graduates. If critical, ethical, and productive engagement with AI is taught and integrated into assessment tasks in meaningful ways then students will regard it as an essential part of their learning, rather than a supplementary component.

> **Examples:**
>
> A Bachelor of Commerce sets a digital mastery outcome for its graduates and designs a marketing subject on developing appropriate disciplinary capabilities with respect to generative AI. Assessment includes both a task creating appropriate generative AI prompts and a class presentation on the ethical and legal pitfalls of using large language models based on this task.
>
> An assessment in a postgraduate engineering qualification requires students to use AI to design and create a robot. Associated documentation identifies the specific limitations and risks of using and not using AI within specific design choices.

2. ...a systemic approach to program assessment aligned with disciplines/ qualifications

Assessment operates across a program of study. Assessment design considerations should span a whole program and/or the sub-structures within, such as majors, rather than be applied solely at individual task or unit levels. Assessment then becomes a matter of educational design, allowing for multiple methods, integrated tasks, and meaningful feedback/dialogue between educators and students to support judgements about progress and attainment.

A systemic/programmatic approach to assessment provides multiple means for educators to make judgements about student progress, without losing the emphasis on feedback and dialogue. These judgements can be captured or tracked over time as student knowledge and skills develop. This in turn promotes the trustworthiness of the overall award rather than relying on a series of singular, uncoordinated judgements. This type of approach becomes a core consideration in ensuring appropriate credentialing due to the emergence of generative AI.

> **Examples:**
>
> A physiotherapy degree maps out all its assessments to a series of clear graduate learning outcomes, that are supported by many small, graded tasks, including written tasks, engagement with clinical technologies including AI, observed clinical skills, interactive orals and supervisor reports on clinical placements. These tasks are collected, collated, and analysed periodically using a portfolio (with a coach/mentor) to monitor attainment and progress towards a decision point.
>
> A Bachelor of Arts develops an approach to assessment that spans the program of study but also allows cohesive majors within particular disciplines. Overall learning outcomes are developed for disciplinary specialisations and for the program overall and include an outcome about digital mastery. Students experience a series of tutorials with many interactive orals integrated across key themes, becoming increasingly in-depth as the student majors in one or two disciplines. Tasks are explicitly aligned with program and disciplinary learning outcomes.

3. ...the process of learning

Evidencing the process of learning over time and in context can support a better understanding of learners' sense-making processes, what they ultimately know and can do. Learning tasks should provide opportunities to reveal thinking, competencies and other qualities embodied in learning outcomes. Assessment tasks should provide opportunities for feedback on the products of student work (e.g. an essay, a laboratory report, an industry brief, or a video submission) that reflect critical thinking, judgement, decision-making, including ethical decision-making, and reflection on the process. These are components of the process that AI is less able to simulate.

The assessment of written or otherwise produced pieces of assessable work is a problem in the context of AI because learning processes are inherently opaque to educators and often to learners. This results in difficulty identifying what has been created by humans and what has been created by AI. Where it is appropriate for assessable products to be created by both AI and students, the assessment design should provide clear opportunities to gather evidence where learners critically engage with the use of AI, demonstrate judgement in how to best use AI and reflect on the learning process.

Examples:

In a Bachelor of Science, the physics department seeks to break a large assessment into constituent parts and activities, showing changes over time. The students provide evidence of their critical thinking by revealing their decision-making, including dead ends and how the task relates to relevant practicals and tutorials. The assignment is a collection of evidence of student processes rather than a final product: the focus is on how the product is derived, not on how generative AI simulates its derivation.

In a postgraduate law subject, students undertake an essay arguing a position on a legal issue that iteratively builds on feedback from different sources: peers, self, teacher and generative AI. The explicit inclusion of generative AI requires students to distinguish between the credibility of the sources and to provide a specific rationale for their choices. The submitted assessable product focuses on responding to this feedback and gathering evidence of the development of their judgement when forming a legal argument.

4. ...opportunities for students to work appropriately with each other and AI

Assessment designs and judgements of student achievement should provide the opportunity for good quality collaborative work. This would include an articulation of, and design for, the acceptable ways students work with each other and AI. There is evidence showing that AI is being integrated into a complex network of interactions with machines, peers, and educators; therefore, students working one-to-one with AI is not the only way in which these technologies should be used.

It is critical that students are able to collaborate appropriately with each other, incorporating AI, and with other technological tools. Students should describe and reflect on the role of people and technologies in how the collaborative work was carried out. A necessary condition for trustworthy assessment is articulating to students the boundaries of what is and is not permissible when working with AI. Through this articulation, collaborative practices can be highlighted that will support students to engage with each other and AI. Additionally, to ensure more inclusive assessment, access and/or capabilities for working with AI should not be assumed but incorporated into curricula. Moreover, through this articulation, collaborative practices can be highlighted that will support students to engage with each other and AI.

Examples:

In a journalism subject, the assignment brief articulates that a group of students can use generative AI to help with drafting the written response to the brief, but the students must explain what prompts were used, what ideas were generated and how the group then collaboratively shaped the contribution of the tools used.

A Master of Interior Design has a capstone creative project. Students are asked to work in groups with AI tools to produce a design brief and then assessors make judgements about both the nature and the productivity of the collaboration. As part of this task, students keep a diary or log of the collaborative process that is also assessed.

5. ...security at meaningful points across a program to inform decisions about progression and completion

In many disciplines, there may be a need to understand and evidence what students are capable of without AI. Rather than attempting to secure every assessment task against the use of AI, this proposition involves identifying the key assessment moments at a program level and securing those. These assessment moments are likely to be related to program-level learning outcomes and are important either to the student's journey through the course, or to judging program completion. Importantly, this proposition differs but works in parallel to proposition 2; it complements the holistic notion of systemic assessment, with a focus on specific, critical points within a program.

Because assessment security of the type required to limit the use of AI is often resource-intensive, it is not feasible to apply these approaches to all assessments. There is a need to target this approach to where it will have the most impact – namely those moments of assessment that provide greater assurance that students who have been awarded the qualification have achieved the program outcomes.

> **Examples:**
>
> A Bachelor of Computer Science program team identifies the capstone project as a key moment for assuring learning for the degree. To secure the projects, academics meet with students across the study period and engage in interactive code reviews. To resource this change, the exam is removed from the first-year introductory unit, which is now 100% within-semester assessment.
>
> A Bachelor of Science program team notices that, while they have exams in almost all first- and second-year units, there are none in the third year of the program, which is all unsupervised take-home tasks. This concerns the team, as the final year is tasked with certifying student achievement. The team decides to make a long-term plan to review assessment across the program to see how they could integrate a multi-method assessment approach for key assessment moments while maintaining security. The team remove some exams from the first year and introduce an exam in one final-year unit per major. This results in fewer exams in the course overall, but more targeted assurance of learning.

Conclusion

The guiding principles and propositions outlined here are offered to help inform decision-making and action into the future. However, they are not a formulaic prescription and are agnostic to the mode of delivery. Indeed, the principles and propositions presented here will create different challenges in online and external modes of study. These principles and propositions will likely be applied to those contexts in different ways but apply nonetheless.

The ideas in this document should also be viewed in an integrated manner. An approach to assessment in on-campus, external, online or blended modes would incorporate elements across the principles and propositions rather than rely heavily on one or two. Doing so involves work at the qualification or program level and not just on a unit-to-unit basis.

As these technologies evolve, this document is intended to guide ongoing discussion and decision-making and will be revisited and updated accordingly. These conversations will necessarily involve exploring the relationships between assessment, teaching, delivery mode,

and learning within a complex, changing environment. To accommodate the reforms set out in this document, Higher Education Providers may also need to ask what they need to stop teaching and assessing, as some learning outcomes start to lose relevance or become infeasible to assess. Critically, a range of voices including students, educators, educational leaders, equity practitioners, educational technologists, educational designers, First Nations peoples, peak bodies, technology vendors, and professional accreditation agencies should be included in these discussions.

We recognise that change in assessment practices at the scale required will not be straightforward. There are workload and accreditation matters, among other complex issues that will be barriers to the changes implied in this document. There will be a widespread need to build academic and professional staff capabilities for working with AI. It should be recognised that AI also presents opportunities for enhancing practice, for example in feedback, in addition to the risks we have outlined and responded to in this document. Together, the sector can harness the opportunities offered in this time of change to enhance assessment design and to build a better higher education system emphasising partnerships, consensus, and the importance of knowing and working alongside our students.

Process

This document is the culmination of a collaboration of experts in artificial intelligence, assessment, and higher education in Australia. The document represents the key guiding principles and propositions arrived at through two days of discussion at a forum hosted by TEQSA and University of Technology Sydney in August 2023.

References

Boud, D. and Associates (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Australian Learning and Teaching Council.

Higher Education Standards Framework (Threshold Standards) (Cth). www.legislation.gov.au/Details/F2022C00105.

National Artificial Intelligence Taskforce (2023). *Australian Framework for Generative Artificial Intelligence in Schools*. New South Wales Department of Education.

THIS PAGE HAS BEEN LEFT BLANK INTENTIONALLY

TEQSA