EduGrowth

cradle Centre for Research in Assessment and Digital Learning

inspera

# How do we balance assessment security and flexibility in e-assessment

September 2022

# About EduGrowth

EduGrowth is Australia's education technology and innovation industry hub. Through connection and collaboration we accelerate Australia's EdTech ecosystem globally.

EduGrowth connects a community of education providers, industry participants and EdTech entrepreneurs committed to reimagining learning in the digital age. As education transitions to borderless digital delivery, EduGrowth's diverse ecosystem will impact the future of learning globally from Australia.

# About Inspera

Inspera Assessment was created to provide test-takers with equal opportunities and to enable them to prove their skills in an authentic way. We empower universities, awarding organisations, governments, municipalities, and schools to deliver secure, standardised tests, open and closed book exams, as well as formative and diagnostics tests.

Inspera offers the most innovative, reliable, and secure online assessment platform, accompanied by the necessary expertise to ensure successful organisation-wide implementation and rollout. Founded in 1999, we are a dedicated international team of more than 130 EdTech entrepreneurs on a mission to impact education on a global scale.

# About the Author

**Professor Phillip Dawson** leads research into academic integrity at the Centre for Research in Assessment and Digital Learning (CRADLE), at Deakin University in Melbourne, Australia. He holds degrees in education, artificial intelligence and cybersecurity. Phill leads CRADLE's research agenda on academic integrity, with a focus on experimental studies and new technologies. Phill has published some of the first experimental studies on contract cheating detection and computer-based exam hacking. He is currently engaged in research on different approaches to detect and deter contract cheating, including assessment designs and technologies. He also has a keen interest in how academics make decisions in assessment design. He uses his background in assessment and cybersecurity to protect education from cheating. His work involves unorthodox methods like computer hacking and paying professional cheaters.

Please cite as: Dawson, P. (2022). *How do we balance assessment security and flexibility in e-assessment* [Whitepaper]. EduGrowth and Inspera.

# Introduction

During the COVID-19 pandemic there has been a significant shift towards e-assessment, which has brought with it concerns about student cheating. In the early part of the pandemic, many educators were assessing in 'survival mode' (Meccawy et al., 2021); doing the best they could with the resources available. Assessment needed to move online, fast. But it appears that much of this shift has happened for good – so it's time to tackle the big questions of e-assessment.

The tension between increasing flexibility and securing assessment against cheating has been one of the biggest challenges in this new reality. On the one hand, students need to complete assessed tasks at a location of their choosing, on their own devices, and potentially at a time that suits them – in other words, flexibly. But if we cannot physically observe students while they do these high-stakes tasks, are we inviting cheating?

During the early pandemic period, educators scrambled to deliver flexible assessment using the tools available to them, attempting to do a good enough job to assess students where they were, using approaches they thought would minimise cheating (e.g. Khan et al., 2021). At an education systems level, moderate success was achieved: students were assessed at locations of their choosing (Reedy et al., 2021), but there was a very significant increase in cheating rates, as evidenced by, for example, the substantial increase in the use of multi-billion-dollar cheating companies (Lancaster & Cotarlan, 2021), and increases in self-reported cheating compared to surveys conducted pre-COVID (Curtis et al., 2021).

To explore this challenging terrain, Inspera engaged EduGrowth and Professor Phillip Dawson from the Centre for Research in Assessment and Digital Learning (CRADLE) at Deakin University to analyse the literature and consult with experts and leaders from across the sector. A workshop was held with 17 senior leaders and experts from Australian higher education and the educational technology sector. In that workshop, participants discussed how to balance assessment security and flexibility in e-assessment. This report builds on the research literature, participants' expertise, and examples of good practice drawn from across the sector.

# What is assessment?

Before we consider flexibility and security, it is worth clarifying what exactly we mean by assessment. To assess is to observe what students have done, and from this, make inferences about what they are capable of (Joughin, 2009). Assessment serves multiple purposes: certifying that students have met particular learning outcomes; guiding students through meaningful learning tasks; and developing students' ability to make judgements about the quality of their and others' work (Boud, 2000). Assessment is high-stakes; in some ways it represents the most unavoidable part of a course of study, as it must be engaged with in order to graduate (Boud, 1995).

# Cheating, academic integrity and assessment security

In recent years, cheating has become more sophisticated and diverse. Traditional forms of cheating such as blatant plagiarism and in-person exam cheating are now accompanied by assignment outsourcing or 'contract' cheating; coordinated efforts to pool answers for test questions in real time; and the use of artificial intelligence to produce written tasks. This fuller set of cheating threats is what assessment designers must now contend with. Combined analysis of all available studies on the prevalence of commercialised cheating in 2014-2018 estimated that 15.7% of students had paid someone else to do their assignments (Newton, 2018); work conducted during the pandemic suggests rates have increased in recent years. Workshop participants expressed their frustration at the challenges of addressing cheating:

> "Responding to academic integrity and cheating behaviours is challenging.

Academic integrity is a positive, educative and values-based approach to addressing cheating. The International Centre for Academic Integrity proposes that it consists of six fundamental values: honesty, trust, fairness, respect, responsibility, and courage (Fishman, 2014). A variety of strategies fall under the academic integrity banner: educating students so they can do, and choose to do, the right thing, perhaps through academic integrity modules (Sefcik et al., 2020); asking students to sign on to an honour code saying they will do the right thing (McCabe et al., 2002); and integrating the teaching of relevant skills, such as referencing, into everyday teaching and learning practice.

In keeping with this positive mission, academic integrity does not include approaches that monitor student compliance with the rules and/or attempt to make breaking the rules more difficult. These more adversarial approaches fall under the banner of 'assessment security', and include remote proctoring, randomizing test questions, and rewriting assessments from year to year (Dawson, 2021). If an approach to addressing cheating focuses on surveilling students, hardening tasks against cheating, or analysing student responses to detect potential cheating, it is an assessment security approach, not an academic integrity approach.

Academic integrity and assessment security are both essential to address cheating. Neither is sufficient on its own. They are akin to modern approaches to addressing crime, which incorporate both crime prevention and policing. The public entrusts educational institutions with a mission to graduate students they are sure have met the learning outcomes and are ethical people.

# Assessment, student motivations and cheating

Beyond our needs for assessment to be secure, and for students to act with integrity, we also need to consider students and what they want. What motivates them? Ramsden (1992) claims that assessment "always defines the actual curriculum" from the student perspective (p. 187). Assessment therefore plays a key role in student motivation, but this is not always positive. The evidence that assessment can necessarily drive students towards learning is debatable (Joughin, 2010), and workshop participants commented that some assessment might instead drive cheating:

> Suppose a student has weighed the importance of the assessment task equal to a game of cards or monopoly. In that case, integrity and cheating aren't something they are going to take too seriously. Your definition of "doing the right thing" may not be aligned with a student's definition. It is very situational. For example: Assisting my friend and classmate is more important to me than following your codes.

Two fruitful directions for considering student motivation and cheating were discussed in the workshop. One of these was considering motivational theory, particularly Self Determination Theory (Ryan & Deci, 2000), which focuses on intrinsic and extrinsic motivation. This specific language of intrinsic and extrinsic motivation was used by workshop participants:

> The nature of students' extrinsic and intrinsic motivation (and the notion of self-determination) are key to cheating in assessments (and other activities).. so we need to better frame this for students.

Substantial evidence on SDT finds that intrinsic motivation is supported when people have a sense of autonomy, a sense of competence, and a sense of relatedness. Assessment design that supports these needs is likely to fuel intrinsic motivation and reduce cheating. For example, being forced to do a task you do not want to do is likely to threaten a student's sense of autonomy. This brings us to the second fruitful direction that was discussed: asking students what types of tasks they would be more or less likely to cheat on. Figure 1 reports data from a large scale survey of more than 14,000 Australian students (Bretag et al., 2019) when they were asked what types of tasks they would be more or less likely to contract cheat on, a type of cheating that involves getting someone else to do the work for you.

## Students' perceptions of the likelihood of contract cheating (%)

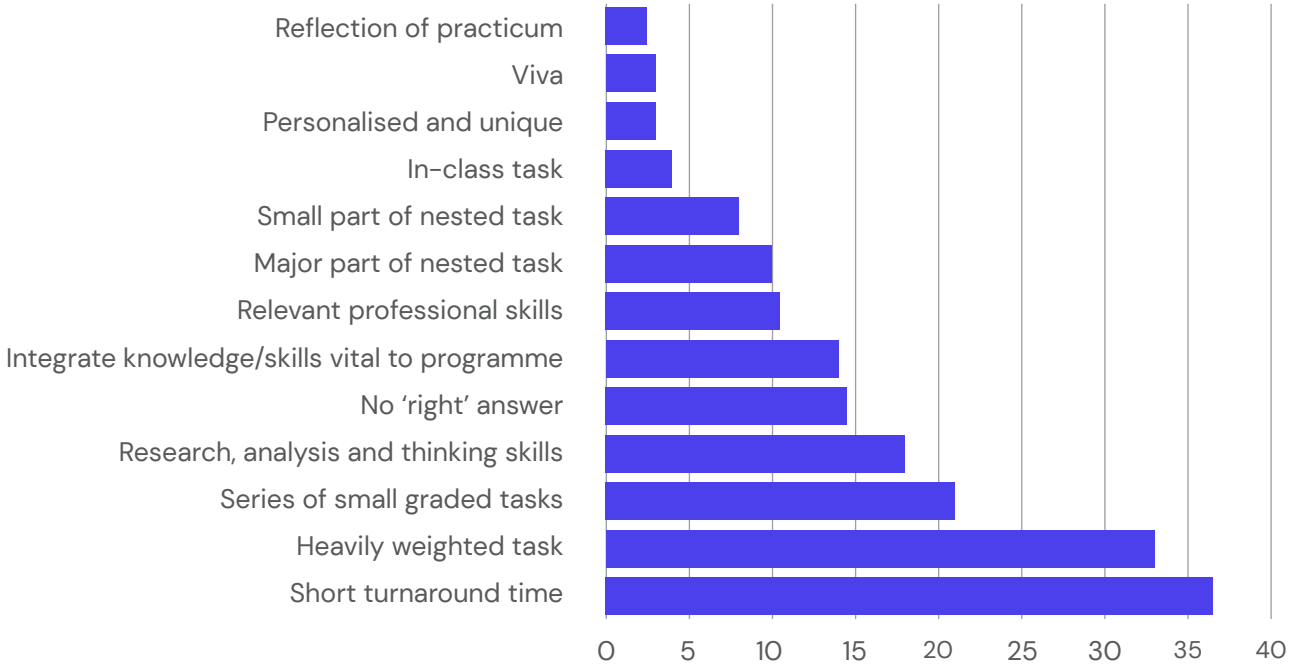


Figure 1: Proportion of students who gave a 'likely' or 'very likely' response when asked about the likelihood of contract cheating on different types of task. Data from Bretag et al. (2019).

This data reveals that some types of tasks are probably more likely to be cheated on than others. Tasks that rate the lowest in terms of likelihood to contract cheat are tasks higher in relatedness, in that they involve interactions with other people or connecting to previous interactions with other people. The task characteristic rated as most likely to be contract cheated on, tasks with short turnaround time, impinges on personal autonomy by restricting the time that can be taken. Students want assessment that is more flexible – but what does this mean in an online context?

# Flexibility in online assessment

What does it mean for an assessment to be flexible? At a minimum, there must be an element of choice for students (Irwin & Hepplestone, 2012). This usually plays out in terms of choice of location, time or aspects of the task itself such as a choice of questions. An exam might, for example, be made more flexible by offering it online on a device of the student's choosing rather than face-to-face. The literature also contains other understandings of flexibility, such as the option for students to select and/or weight tasks in different ways as best suits them (Cook, 2001; Rideout, 2018).

Workshop participants were student-focused in their considerations of flexibility. They noted that while flexibility can be desirable for students, and the efforts put in by staff to support flexibility can be very substantial, flexibility can also come at a cost for students:

> "Allowing students to undertake exams on their own devices and in preferred locations provides more flexibility. We also need to realise that we have shifted the responsibility to them. A student is now responsible for the technology, WiFi and appropriateness of the location and exam setting. A proportion of students don't want this responsibility.

Despite this, one workshop participant noted that "Students can be more flexible with changes in assessments and platforms than staff are." They thought that the flexibility of students in assessment should be appreciated, especially over the recent pandemic period.

For assessment to meet its summative purpose, flexibility must not impact negatively on the validity of the task (Irwin & Hepplestone, 2012). Validity is a complex concept, however, underlying most modern conceptualisations of validity is a concern that an act of assessment assesses the learning outcomes it is supposed to assess, and does not unintentionally assess other spurious constructs (St-Onge et al., 2017). Validity and flexibility are intertwined. Taking the flexibility approach of giving students a choice of tasks, if validity is to be maintained, each of these choices should be equally good at allowing assessors to judge if students have met the learning outcomes. However, the inverse is also true: for inflexibility in assessment to be justifiable, the case must be made that this inflexibility does not hurt validity. Traditional face-to-face exams, for example, can be criticised as assessing a student's ability to get to a particular exam venue at a particular time, and deal with exam anxiety, alongside assessing whatever outcomes are officially being assessed.

# Resolving the tension between flexibility and assessment security

Flexibility and assessment security are in tension with each other. Providing students with options means that some of these options might be more vulnerable to cheating. But conversely, privileging assessment security can mean denying students the flexibility they need to fit their assessments into their lives.

There are several ways to resolve this tension, and this section covers two ways of thinking about it. The first is to consider assessments in terms of their flexibility and assessment security, and to look for tasks which are higher in each. Figure 2 proposes how this might be done, by placing assessments into quadrants. The most desirable quadrant is the top-right one, where tasks are both high assessment security and high flexibility.

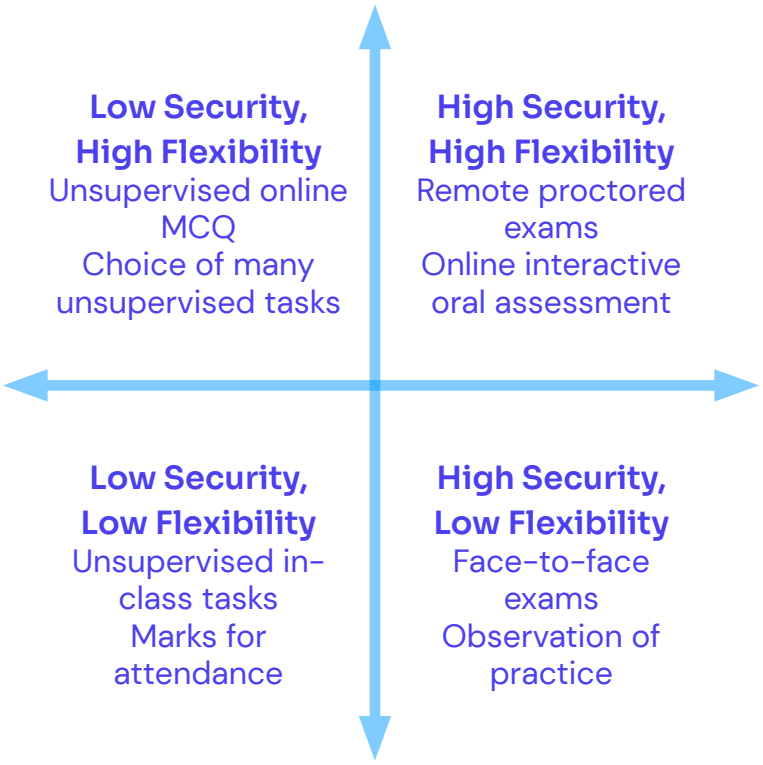


Figure 2: Four quadrants of assessment security and flexibility in e-assessment.

Workshop participants provided specific examples of top-right quadrant assessments during the workshop. Three key examples were:

1. Interactive Oral Assessment by Danielle Logan-Fleming, Popi Sotiriadou, Amanda Daly and Ross Guest (Griffith). They describe an Interactive Oral as "not a question and answer test, but rather an exchange which draws upon the student's understanding, and creates a setting in which they can demonstrate and apply course concepts"
2. Collaborative online exams by David Kellermann (UNSW). These exams go beyond traditional open book to also involve collaboration between students via a Microsoft Teams site, where students ask questions and help each other. Rather than attempting to restrict student collaboration, this type of exam is designed with collaboration in mind.
3. Virtual work integrated learning by Sally Male (formerly UWA, now Melbourne), in which "students undertake learning activities that involve industry but are not true employment (paid or unpaid). Students complete authentic tasks, using authentic tools and/or processes, and engage face-to-face or electronically with real or simulated workplaces and/or practitioners"

A challenge with assessments in the top-right quadrant is that they can be more expensive to operate. As with all assessment design decisions, the choice to deploy these assessments should be made programmatically. Programmatic assessment involves looking at assessment at the level of the qualification, not just at individual acts of assessment, and ensuring that all of the program-level outcomes are assessed (van der Vleuten et al., 2012). From an assessment security perspective this means ensuring that high-assessment-security approaches (those in the two quadrants on the right) are deployed at key summative moments throughout the qualification. There may be only a handful of such moments across a program of study. For other moments of assessment, which should make up the majority, it is less important to privilege assessment security, so assessment types that are more flexible, efficient, or in some other ways suitable, can be used instead. There was support among the workshop participants for this idea, but also recognition that it does come with challenges:

> “Linked assessments build a lot of security but also act as supporting scaffolds for students. The big issue to overcome is the atomisation caused by courses not linked to each other.

‘Atomisation’ here refers to the breaking down of degrees into course units, and often beyond that into smaller pieces such as modules within units, or microcredentials. While this approach can provide greater flexibility in terms of being able to ‘mix and match’ to create many different pathways through programs of study, they make it difficult to view assessment at the level of the degree program.

The second way of resolving the tension between flexibility and assessment security is to think of them as ways of enhancing validity. Cheating threatens validity by invalidating assessor judgements; you can’t judge what a student is capable of if they have not done the work in the conditions you have prescribed. Similarly, inflexibility threatens validity by making assessment less inclusive; you can’t judge what a student is capable of if they were unfairly disadvantaged. Validity is the most important component of assessment design (American Educational Research Association et al., 2014), and improvements to either assessment security or flexibility are improvements to validity.

# Ten considerations for balancing assessment security and flexibility in e-assessment

## 1 Authenticity is not a panacea, but a sound argument needs to be made for inauthenticity

Authentic assessment has a variety of meanings, but common to most is a desire for realism in assessment, through setting students problems similar to those they will encounter when they graduate, in contexts that represent the real world of their chosen discipline (Villarroel et al., 2018). In and of itself, authentic assessment does not stop cheating; for example, commercialised cheating services can and do produce vast quantities of authentic tasks (Ellis et al., 2019). That said, authentic assessment can offer a useful razor that can enhance flexibility and assessment security: the setting of authentic restrictions (Dawson, 2021). By looking at the tools, information and people a professional would usually have access to when undertaking a task, an assessment designer can see if the restrictions imposed on students when undertaking that task as an assessment are authentic. For example, when undertaking a medical diagnosis task, a doctor typically has access to a range of texts, as well as access to colleagues. Denying a medical student the flexibility to access those same resources should only be done if allowing them access would get in the way of assessing the learning outcomes. Because, as one participant noted:

> Everyday we rely on apps and online for information, why do we force learners to know EVERYTHING with information readily available. Learners need to be skill ready above all.

Restrictions make assessment less flexible, and, somewhat unintuitively, restrictions on assessment can also hurt assessment security rather than enhance it. For example, setting an exam as a closed-book task means that this restriction must be enforced. If a subset of students successfully access their textbooks during a closed-book exam then the validity of the assessor judgement for those students is threatened; the task is also made unfair for those students who do not use their books. Every restriction needs to be enforced, and every restriction increases the 'attack surface' that needs to be defended for a task (to borrow a term from cybersecurity). Restrictions should therefore only be in place where they are absolutely necessary, for example, if a task focuses on lower-level learning outcomes that can be essentially looked up in a book. For higher level outcomes, assessment designers should ask: what rationale do I have for any restrictions that are not authentic?

## 2 Scale favours frontloading of assessment design efforts

Educators are increasingly expected to do more with less when it comes to assessment. Excellent e-assessment practices are still possible at large scale within traditional workload models – but they require significant investment in design (e.g. Broadbent et al., 2018). Assessment designers should consider which approaches scale linearly (e.g. double the students takes approximately double the work, as is the case when marking essays), and which approaches scale better than linearly (e.g. double the students requires little extra work, as is the case with multiple choice questions, or high fidelity simulations) (Dawson & Bearman, 2020). As one participant noted, this should be considered when apportioning time and effort in the assessment process:

> The time allocations and workload counting should come at the start to load the design and development of the assessment. This should drive the assessment throughout the course and not require such 'hour counting' at the end-point.

The larger the cohort, the more important it is to 'frontload' assessment efforts into better design, which can enable flexibility and assessment security.

## 3 Shelf-life and sustainability need careful consideration

For assessment types with a finite range of answers – such as multiple choice – once a question has been exposed to a group of students it is, to an extent, used up. Thanks to sites like Chegg, Course Hero, and Quizlet, there are now anonymous large-scale commercial ways for students to share test questions and answers in addition to the traditional informal ways of sharing them that have long existed. Reusing questions and/or using the same test for a large group of online test takers is to an extent just inviting cheating. This problem is greater if test questions were exposed to learners in a low-security context, such as an unsupervised online test. Where some sort of supervision has been used, there is a smaller but still significant chance of questions being leaked.

This new reality of question sharing places greater importance on smarter and more adaptive tests. A single bank of questions that all students take is not enough. For questions to have any shelf life beyond an initial use, several techniques should be used. Workshop participants echoed the suggestion of Elkhatat (2022), who finds that a bank of items several times as large as the number any student will sit is required, or even multiple banks of items. Custom items that incorporate randomness or participant characteristics were also discussed, such as using different numbers in a calculation exercise. The use of QR codes as digital watermarks was also discussed as being used in some online exams, as a way to identify when they have been uploaded to cheating websites. Future technologies capable of creating unique items based on underlying learning outcomes should be pursued.

## 4 Build for the future of assessment security and flexibility

The pandemic demonstrated that sometimes rapid changes need to happen to enable different types of flexibility in e-assessment and different standards for assessment security. But this is unlikely to be the last time rapid changes are needed. Participants noted that most of our thinking and examples in this space is now at least two years old. Writing and image development tools such as GPT-3 and DALL-E which can produce human-quality work for student assessment, indicate that there is a need to reconsider the role of AI in assessment. This has bearings for both assessment security – how can we be sure students have met the outcomes themselves – and for flexibility – how can we decide what tools students are to use?

At all levels, from local assessment regimes through to federal legislation, policy and procedure should take a principles-based approach rather than referring to the specifics of current situations and technologies. With one of the goals of the artificial intelligence movement being passing the 'robot college student Turing test' – building an AI that can fulfil the requirements of a degree (Goertzel et al., 2012) – the threats to assessment security may change rapidly and unpredictably. And with other societal threats such as climate change and students in conflict zones, flexibility needs to be built into policy in ways that may not be immediately obvious. For assessment to operate in this new world, flexibility needs to be the default, not an optional add-on.

## 5 Assessment security and flexibility disproportionately affect some groups of students

Whatever is done to reconcile the tension between assessment security and flexibility, needs to be done with diversity and inclusion in mind. E-assessment has enabled many students to participate in high-stakes assessment from their own homes, and this should be celebrated. But there is a need to consider if the flexibility afforded to some students is afforded to all students. Students who have registered disabilities are usually provided adjustments to enable them to undertake some types of assessments such as exams, but best practice is to redesign assessment so that students can undertake assessments without adjustments (Tai et al., 2022). In this context, flexibility means not having to ask for adjustments. Similarly, assessment security needs to be carefully scrutinised, as some groups of students are persistently overrepresented in cheating statistics. Workshop participants noted in particular that students who face financial difficulties, language barriers or are international students may have particular challenges in undertaking the assessments that are set for them. The conditions that lead to their cheating, as well as potential biases in assessment security technologies, need to be addressed (Eaton, 2022).

## 6 Assessment change is a team sport

Whatever is done to address the challenges of flexibility and assessment security, if it involves changes to assessment then it will involve many different people in different roles; as one participant noted, "No one individual can innovate in education". The specific roles mentioned by workshop participants included academics, learning designers, media producers, and exams teams. Senior university leadership were also regarded as having a role in terms of signalling that assessment design is something worth investing in.

Assessment design was regarded as an expert practice. One participant went as far as to suggest that academics should be removed from the assessment design process. While this was not a consensus opinion, there was fruitful discussion around the expertise that learning designers bring to assessment design, and the emergence of specialist learning designers who just do assessment work. Structures to build assessment expertise and innovation were also discussed, with the University of South Australia's Assessment Design Lab and Deakin's Centre for Research in Assessment and Digital Learning (CRADLE) being presented as exemplars.

## 7 Regulators and accreditors are both influential and influenceable

Regulatory bodies such as the Tertiary Education Quality and Standards Agency (TEQSA) as well as professional accrediting bodies were regarded as powerful players in assessment. Much of what was done in assessment, particularly around assessment security, was done to address concerns from these bodies; it was noted by one workshop participant that:

> "Accreditation requirements are pushing innovation, Accreditation needs eased up over Covid but has since been reintroduced. There are questions around authenticity in design for remote examination while meeting industry needs.

This speaks to a perception in the workshop that regulators and accreditors privileged assessment security over other aspects of assessment, leading to trade-offs in terms of authenticity. However, there was also a sense that these bodies can or should be influenced, especially when institutions, disciplines or professions lobby together:

> Educators should push back on industry, but colleagues often do this in isolation.

> Restrictions around assessment come from industry / accreditation bodies assessment. Need a collective push back or challenge - to see if they are actually getting the graduates they need.

The prospect of co-design with professional accreditation bodies was raised as a fruitful way forward, both because it may lead to more mutually acceptable designs, but also because this sort of partnership was seen as a way to assure external validity.

## 8 Platforms shape assessment

All of the proposals discussed around assessment that was more flexible and/or secure depended on platforms. There was a strong belief that these tools mattered, and that they shape what happens in assessment:

> The affordances of technology impacts the shape of assessment. We need to select / create tech carefully.

This opinion is backed up in the literature, such as in a study by Bennett et al. (2017) of Australian university educators, which found that technology affordances powerfully influenced assessment design. Participants also noted that this shaping was not unidirectional: "We shape our tools and they shape us."

Over the course of the workshop, participants discussed a broad range of desirable affordances: video calls, media content, online document editing, discussion, collaboration, video recorded evidence from students and randomised questioning; the breadth of items on this list demonstrate a desire for an e-assessment ecosystem rather than just a single type of tool. Such an ecosystem comes at a financial cost, and there was concern for the opportunity cost of this sort of investment:

> Unfortunately sometimes the investment in IT infrastructure means there is a lack of investment in other areas such as innovation in teaching.

However, similar concerns were raised about the much more significant expenditures being made on physical learning and teaching spaces that are now underutilised.

There was also nuanced conversation about the different discourses that exist around e-assessment platforms, with one participant recommending the two categories posed by Allan (2020) as a useful way to conceptualise how we think about e-assessment. In the context of e-exams, Allan identified two discourses: migration, in which e-exams were seen as "neutral instruments used independently by humans to realise their preordained intentions"; and transformation, in which "the essential and inalienable qualities of technologies can be released to 'transform' or 'enhance' assessment".

## 9 Universal challenges of teaching and learning in higher education matter in assessment too

Much of what was said by workshop participants about the challenges of addressing flexibility and security in e-assessment has been said many times before, about many other learning and teaching problems, for example: the prioritisation of research over teaching, the casualisation of the university teaching and learning workforce, and workload models that do not reflect the time taken to do teaching and learning work. These sorts of concerns often come up as barriers when assessment redesign is discussed (Bearman et al., 2017; Deneen & Boud, 2013), as well as when other learning and teaching problems are discussed, and they are familiar to most working in this space. However, it is important to mention them here, as they pose a real threat to the viability of all other suggestions in this document.

The degree to which these problems are intractable was up for debate; while government funding was identified as a root cause for some of these problems, what institutions did with that funding varied wildly. For example, some participants' universities allocated one hour per student per course unit for marking and feedback, whereas one participant from a comparable institution said their university allocated twice as much time. However, they said that this doubled per-student assessment resourcing was not a panacea, and the difficulties they faced in assessment were similar to those across the sector. Whatever work is done in this space needs to connect to these realities of working in teaching and learning in higher education.

## 10 We need to acknowledge the problem of cheating and share intelligence

This document opened with a discussion about cheating at universities. Some participants regarded this as a problem that universities do not wish to admit to, especially publicly. One participant mentioned an Australian university that did acknowledge rates of cheating publicly, and the negative reporting that resulted in the media, concluding that "it's a difficult problem for universities to admit to." However, if the problem of cheating is to be addressed, institutions need the capacity to benchmark against each other in terms of the effectiveness of their approaches to promote academic integrity and assure assessment security. Closed workshops, such as the one that led to this report, may be one way forward in discussing the relative effectiveness of different approaches to addressing cheating, and sharing intelligence about the emerging threats on the horizon.

# Acknowledgements

This document would not have been possible without generous input from the following senior leaders and experts from across Australia's higher education and educational technology sectors: Alex Steel, Director of Teaching Strategy at UNSW; Antoine Goarin, Innovation Lead Program Manager, Student Administration Services at Deakin University; Denise Wood, Sub-Dean Academic Development, Division of Learning and Teaching at Charles Sturt University; Gerard Rayner, Associate Director of the Learning Transformation Unit at Swinburne University; Jesse Watts, Exam Delivery Manager at Monash University; Kris Ryan, PVC (Academic) at Monash University; Mathew Hillier, e-assessment Academic at Macquarie University; Phill Dawson, Professor & Associate Director, (CRADLE) at Deakin University; Rebecca Awdry, Researcher in educational quality at Deakin University; Sandy Barker, Senior Lecturer UniSA Business at University South Australia; Pauline Farrell, Managing Director at Skills4theFuture; Caroline Hartnett, Deputy Director, International Education at Global Victoria, Chris Campbell, President at ASCILITE; David Linke, Managing Director at EduGrowth; Luke Seacombe, Head of Engagement at EduGrowth; Sarah Lamont, Programs Coordinator at EduGrowth and Paul Dourlay, Sales Manager ANZ at Inspera. While their input has helped shape the document, the final document represents a best effort at synthesising this input by one author, and does not necessarily reflect the views of each individual workshop participant.

# References


Allan, S. (2020). Migration and transformation: a sociomaterial analysis of practitioners' experiences with online exams. *Research in Learning Technology,* 28(0). https://doi.org/10.25304/rlt.v28.2279

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Bearman, M., Dawson, P., Bennett, S., Hall, M., Molloy, E., Boud, D., & Joughin, G. (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

Bennett, S., Dawson, P., Bearman, M., Molloy, E., & Boud, D. (2017). How technology shapes assessment design: Findings from a study of university teachers. *British Journal of Educational Technology,* 48(2), 672-682. https://doi.org/10.1111/bjet.12439

Boud, D. (1995). Assessment and learning: contradictory or complementary. In P. Knight (Ed.), *Assessment for Learning in Higher Education* (pp. 35-48). Kogan Page.

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

Bretag, T., Harper, R., Burton, M., Ellis, C., Newton, P., van Haeringen, K., Saddiqui, S., & Rozenberg, P. (2019). Contract cheating and assessment design: exploring the relationship. *Assessment & Evaluation in Higher Education,* 44(5), 676-691. https://doi.org/10.1080/02602938.2018.1527892

Broadbent, J., Panadero, E., & Boud, D. (2018). Implementing summative assessment with a formative flavour: a case study in a large class. *Assessment & Evaluation in Higher Education,* 43(2), 307-322. https://doi.org/10.1080/02602938.2017.1343455

Cook, A. (2001). Assessing the Use of Flexible Assessment. *Assessment & Evaluation in Higher Education,* 26(6), 539-549. https://doi.org/10.1080/02602930120093878

Curtis, G. J., McNeill, M., Slade, C., Tremayne, K., Harper, R., Rundle, K., & Greenaway, R. (2021). Moving beyond self-reports to estimate the prevalence of commercial contract cheating: an Australian study. *Studies in Higher Education*, 1-13. https://doi.org/10.1080/03075079.2021.1972093

Dawson, P. (2021). *Defending Assessment Security in a Digital World*. Routledge.

Dawson, P., & Bearman, M. (2020). Concluding Comments: Reimagining University Assessment in a Digital World. In M. Bearman, P. Dawson, R. Ajjawi, J. Tai, & D. Boud (Eds.), *Re-imagining University Assessment in a Digital World* (pp. 291-296). Springer International Publishing. https://doi.org/10.1007/978-3-030-41956-1_20

Deneen, C., & Boud, D. (2013). Patterns of resistance in managing assessment change. *Assessment & Evaluation in Higher Education*, 1-15. https://doi.org/10.1080/02602938.2013.859654

Eaton, S. E. (2022). New priorities for academic integrity: equity, diversity, inclusion, decolonization and Indigenization. *International Journal for Educational Integrity,* 18(1), 10. https://doi.org/10.1007/s40979-022-00105-0

Elkhatat, A. M. (2022). Practical randomly selected question exam design to address replicated and sequential questions in online examinations. *International Journal for Educational Integrity,* 18(1), 8. https://doi.org/10.1007/s40979-022-00103-2

Ellis, C., van Haeringen, K., Harper, R., Bretag, T., Zucker, I., McBride, S., Rozenberg, P., Newton, P., & Saddiqui, S. (2019). Does authentic assessment assure academic integrity? Evidence from contract cheating data. *Higher Education Research & Development,* 39(3), 454-469. https://doi.org/10.1080/07294360.2019.1680956

Fishman, T. (2014). *The fundamental values of academic integrity* (2nd edition ed.). International Center for Academic Integrity, Clemson University.

Goertzel, B., Iklé, M., & Wigmore, J. (2012). The Architecture of Human-Like General Intelligence. In P. Wang & B. Goertzel (Eds.), *Theoretical Foundations of Artificial General Intelligence* (pp. 123-144). Atlantis Press. https://link.springer.com/chapter/10.2991/978-94-91216-62-6_8

Irwin, B., & Hepplestone, S. (2012). Examining increased flexibility in assessment formats. *Assessment & Evaluation in Higher Education,* 37(7), 773-785. https://doi.org/10.1080/02602938.2011.573842

Joughin, G. (2009). Assessment, Learning and Judgement in Higher Education: A Critical Review. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 13-27). Springer Netherlands. https://doi.org/10.1007/978-1-4020-8905-3_2

Joughin, G. (2010). The hidden curriculum revisited: a critical review of research into the influence of summative assessment on learning. *Assessment & Evaluation in Higher Education*, 35(3), 335-345. https://doi.org/10.1080/02602930903221493

Khan, Z. R., Sivasubramaniam, S., Anand, P., & Hysaj, A. (2021). 'e'-thinking teaching and assessment to uphold academic integrity: lessons learned from emergency distance learning. *International Journal for Educational Integrity,* 17(1), 17. https://doi.org/10.1007/s40979-021-00079-5

Lancaster, T., & Cotarlan, C. (2021). Contract cheating by STEM students through a file sharing website: a Covid-19 pandemic perspective. *International Journal for Educational Integrity*, 17(1), 3. https://doi.org/10.1007/s40979-021-00070-0

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2002). Honor Codes and Other Contextual Influences on Academic Integrity: A Replication and Extension to Modified Honor Code Settings. *Research in Higher Education,* 43(3), 357-378. https://doi.org/10.1023/A:1014893102151

Meccawy, Z., Meccawy, M., & Alsobhi, A. (2021). Assessment in 'survival mode': student and faculty perceptions of online assessment practices in HE during Covid-19 pandemic. *International Journal for Educational Integrity*, 17(1), 16. https://doi.org/10.1007/s40979-021-00083-9

Newton, P. M. (2018). How Common Is Commercial Contract Cheating in Higher Education and Is It Increasing? A Systematic Review [Systematic Review]. *Frontiers in Education*, 3(67). https://doi.org/10.3389/feduc.2018.00067

Ramsden, P. (1992). *Learning to teach in higher education*. Routledge.

Reedy, A., Pfitzner, D., Rook, L., & Ellis, L. (2021). Responding to the COVID-19 emergency: student and academic staff perceptions of academic integrity in the transition to online exams at three Australian universities. *International Journal for Educational Integrity*, 17(1), 9. https://doi.org/10.1007/s40979-021-00075-9

Rideout, C. A. (2018). Students' choices and achievement in large undergraduate classes using a novel flexible assessment approach. *Assessment & Evaluation in Higher Education,* 43(1), 68-78. https://doi.org/10.1080/02602938.2017.1294144

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1), 68-78. https://doi.org/10.1037/0003-066X.55.1.68

Sefcik, L., Striepe, M., & Yorke, J. (2020). Mapping the landscape of academic integrity education programs: what approaches are effective? *Assessment & Evaluation in Higher Education,* 45(1), 30-43. https://doi.org/10.1080/02602938.2019.1604942

St-Onge, C., Young, M., Eva, K. W., & Hodges, B. (2017). Validity: one word with a plurality of meanings. *Advances in Health Sciences Education,* 22(4), 853-867. https://doi.org/10.1007/s10459-016-9716-3

Tai, J., Mahoney, P., Ajjawi, R., Bearman, M., Dargusch, J., Dracup, M., & Harris, L. (2022). How are examinations inclusive for students with disabilities in higher education? A sociomaterial analysis. *Assessment & Evaluation in Higher Education,* 1-13. https://doi.org/10.1080/02602938.2022.2077910

van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205-214. https://doi.org/10.3109/0142159X.2012.652239

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education,* 43(5), 840-854. https://doi.org/10.1080/02602938.2017.1412396




Level 12, Tower 2
727 Collin Street
MELBOURNE VIC 3008
edugrowth.org.au

Deakin University
Level 12, Tower 2
727 Collin Street
MELBOURNE VIC 3008
deakin.edu.au/cradle

310 Edward Street
BRISBANE QLD 4000
inspera.com