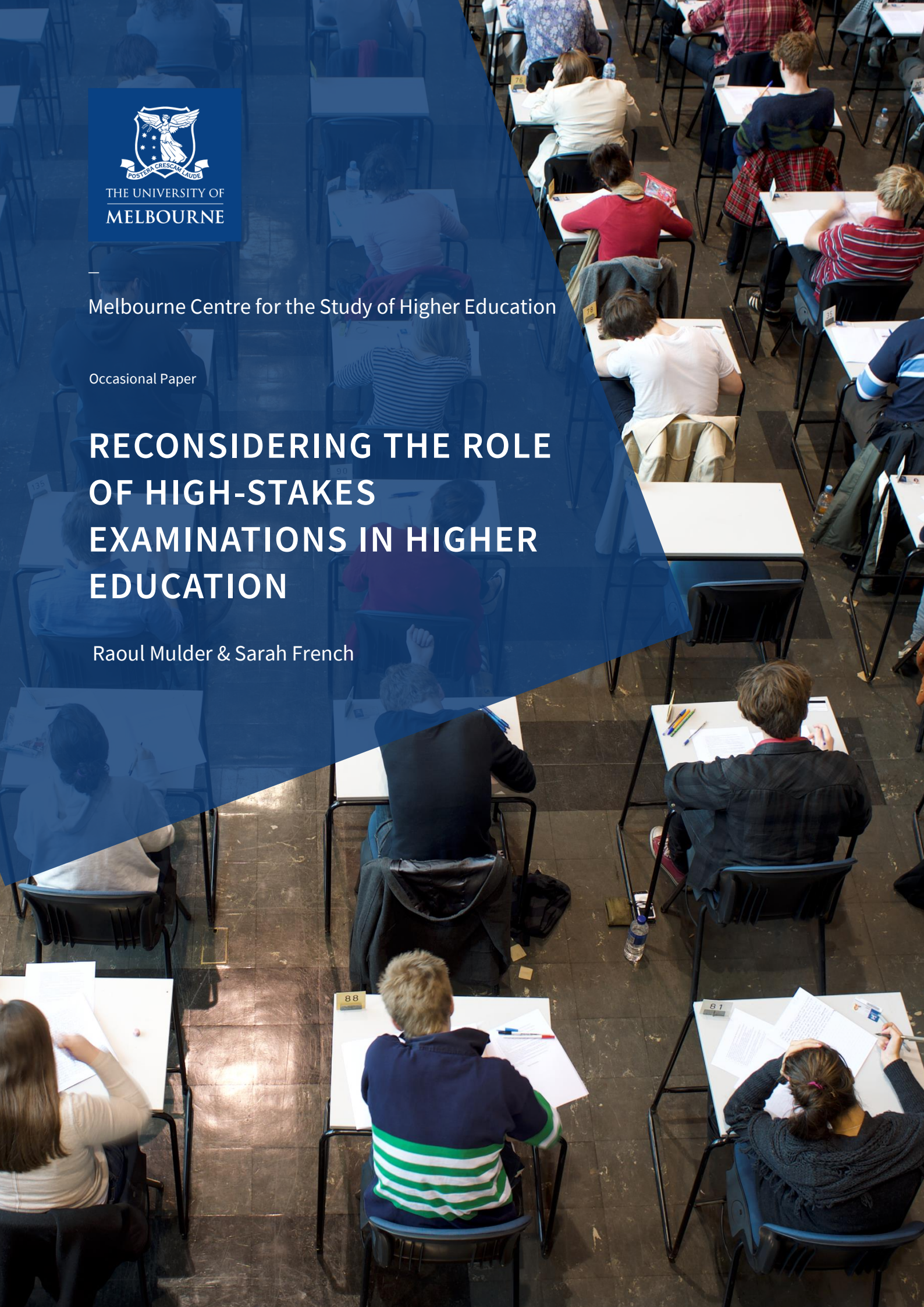

THE UNIVERSITY OF MELBOURNE

Melbourne Centre for the Study of Higher Education

Occasional Paper

RECONSIDERING THE ROLE OF HIGH-STAKES EXAMINATIONS IN HIGHER EDUCATION

Raoul Mulder & Sarah French

Raoul Mulder is Professor of Evolutional Ecology at the School of BioSciences and Professor of Higher Education at the Centre for the Study of Higher Education at the University of Melbourne.

Sarah French is Lecturer in Higher Education at the Centre for the Study of Higher Education at the University of Melbourne.

Acknowledgements: We thank Ashton Dickerson for valuable research assistance, Chi Baik and Sandra Milligan for helpful feedback on an earlier draft of this paper, and the University of Melbourne for funding that supported this research.

Publisher: Melbourne Centre for the Study of Higher Education

Occasional Paper title: Reconsidering the role of high-stakes examinations in higher education

Authors: Raoul Mulder & Sarah French

DOI: https://doi.org/10.26188/21951287

Year of publication: 2023

EXECUTIVE SUMMARY

High-stakes final examinations enjoy widespread use as summative assessments in higher education. In this occasional paper, we review arguments for and against the use of high stakes examinations and evaluate the evidence underpinning these arguments.

From a logistical and resource perspective, high stakes end-of-semester examinations are cost-effective to administer. However, considerations of scale and administrative efficiency must be carefully weighed against the pedagogical merits of such assessments.

The format of high stakes examinations is diverse and varies across academic disciplines and subjects, ranging from closed-book invigilated exams, to open-book, take-home or oral exams, which may be taken in person or online. Nonetheless, most examinations require students to complete the assessment within a constrained window of time and/or with limited access to information. Examinations also tend to be weighted more heavily than other assessment tasks, and therefore often have high-stakes consequences for students.

Perceived academic benefits of high-stakes examinations include their ability to test knowledge recall, motivate study, demonstrate performance under pressure in the absence of information, and their capacity to minimise cheating. Some of these benefits (e.g. knowledge recall, performance under pressure) are arguably authentic only for a very limited set of specialised academic disciplines. Others (e.g. motivation for study) can be achieved equally or more effectively using other forms of assessment. The belief that examinations deter students from cheating is contradicted by abundant empirical evidence.

High-stakes summative examinations have known pedagogical drawbacks. These include concerns about their reliability and validity, their lack of real-world authenticity and relevance, their limited capacity to assess deep learning, their contribution to elevated psychological distress, and their potential to generate academic inequity as a result of gender- or ethnicity-based differences in performance. There is a substantial literature supporting many of these concerns.

On balance, there is much stronger support in the literature for the pedagogical drawbacks of high stakes examinations than for their perceived benefits. Moreover, examinations are poorly suited to assessing diverse learning outcomes and generic skills, and reliance on high-stakes examinations limits the opportunity to diversify the assessment regime.

We conclude that the use of high-stakes examinations in university curricula should be carefully reviewed and provide some options to retain, redesign, reweight or replace the high-stakes examination as an assessment item.

Introduction

Assessment plays a crucial role in higher education, providing a means to certify levels of achievement, verify learning outcomes and test student knowledge. However, assessment has considerable additional educational value due to its potential to motivate, facilitate and enhance learning (Entwistle and Entwistle 1991; Marton and Säljö 1997; Ramsden 1997; Kickert et al. 2022; Carless et al. 2017), as well as lay the foundations for future learning (Boud 2000; 2013; Boud and Falchikov 2006). The way in which students are assessed also has profound implications for both student wellbeing (Slavin, Schindler, and Chibnall 2014; Baik, Larcombe, and Brooker 2019; E. Jones et al. 2021) and student engagement (Vaughan 2014) and, arguably more than any other aspect of teaching, signals to students what is valued by their teachers, the discipline, and the institution.

A good assessment regime must be balanced, fair and unbiased in how it evaluates each student, and provide diversity in assessment tasks to ensure that it can assess a broad range of learning outcomes, maintain student engagement, and involve students in learning activities that lead to higher order thinking and a deeper understanding of content (Biggs and Tang 2007). The importance of constructive alignment between teaching methods, learning objectives and assessment tasks is well established (Biggs 2003), yet there is often a mismatch between stated learning objectives and assessment practices. Assessment is generally designed to prioritise the testing of disciplinary-based knowledge and discipline-related skills at the subject level (Martin and Mahat 2017), often through the use of 'traditional' methods such as essays and examinations. However, there are a range of broader institutional and course-based learning outcomes, aligned with graduate attributes or competencies, that also need to be taught and assessed that are likely to require more varied approaches.

In addition to providing variety in assessment methods, consideration should be given to their role in supporting student learning (Taras 2010; Black et al. 2009), which includes intentional planning around the timing of assessment tasks, and whether they are to fulfil a summative and/or formative function. While summative assessments generally involve the awarding of a grade without feedback or further opportunity to improve, formative assessments give students the opportunity to advance their learning by actively engaging with and implementing feedback (Henderson et al. 2020; Winstone and Carless 2020). Formative assessments can be integrated within teaching and learning activities, offer teaching staff the opportunity to identify areas of weakness and correct them during the teaching process and be scaffolded whereby early formative assessment tasks build towards later summative ones (Boud and Soler 2016). It is widely argued that formative assessment and feedback should feature prominently in curriculum design (Morris, Perry, and Wardle 2021). However, excessive use of continuous formative assessments can result in students feeling stressed and over-assessed (Harland et al. 2015). Therefore, achieving the right balance between summative and formative assessments in the curriculum is key and has significant consequences for student learning.

In addition to calls for more formative and more diversified approaches to assessment, there is growing concern that heavily-weighted summative assessments may negatively impact student learning and wellbeing (Ecclestone 1999; E. Jones et al. 2021; Pascoe, Hetrick, and Parker 2020), and there is debate about their efficacy and validity as assessment instruments (Knight 2002). For all of these reasons, since the 1970s there has been a shift away from the use of high-stakes final examinations in many countries, including New Zealand (Bassey 1971) and the UK (Richardson 2015a), while in many other systems, high-stakes examinations remain strongly entrenched. Researchers in Singapore (Wong, et al., 2020) the U.S (Fook and Sidhu 2014) and Canada (Rawlusyk 2018) suggest that high-stakes examinations continue to be used as a dominant mode of assessment, despite growing awareness of the need for more diversified assessment practices. In Australia there has been a shift away from a heavy reliance on high-stakes examinations (Richardson 2015a), supported by assessment policies at most institutions that limit the weightings that can be attributed to examinations as a percentage of the total grade for a subject. However, many universities, including the University of Melbourne, do not currently have an assessment weighting policy.

Below we examine and evaluate contemporary arguments for the perceived academic benefits and drawbacks of high-stakes summative examinations.

PERCEIVED BENEFITS OF HIGH-STAKES SUMMATIVE EXAMS

EXAMS ENCOURAGE MEMORY RECALL AND KNOWLEDGE RETENTION

There is evidence that testing is a useful technique to improve memory recall and the retention of information. Extensive research on the 'testing effect,' also known as 'test-enhanced learning,' confirms that addressing questions in tests and exams can improve retention (Roediger and Karpicke 2006; Karpicke and Roediger 2008; McDaniel et al. 2007). However, many studies suggest that regular short-answer tests or quizzes taken shortly after the content is taught are of greater value than single, high-stakes summative examinations (McDaniel et al. 2007; Butler and Roediger 2007; Santovena-Casal 2019). Others suggest that examinations undermine long-term knowledge acquisition because they encourage once-off rather than successive relearning (Rawson, Dunlosky, and Sciartelli 2013), and because they emphasise extrinsic reward (Kuhbandner et al. 2016). It is also well known that retention of knowledge demonstrated in exams can be short-lived (H. Jones et al. 2015; Greene 1931).

EXAMS MOTIVATE STUDENTS TO STUDY

Studies examining self-reported levels of motivation have found, perhaps unsurprisingly, that student motivation to study is higher for high-stakes assessments than low-stakes assessments, and that motivation is a positive predictor of outcomes (Wise and Demars 2005; Wise 2009). Study habits are also known to be an important factor in retrieval, retention, and student achievement, including the practices of self-testing, rereading, and scheduling of study (Roediger and Butler 2011; Hartwig and Dunlosky 2012). Such findings suggest that students' capacity to study effectively for an exam may be more important for learning than the sitting of the exam itself. However, if student motivation to study stems from a desire to rote-learn information to perform well on an exam, extrinsic motivation is activated, but not intrinsic or autonomous motivation, which has been shown to be far more important for student learning (R. M. Ryan and Deci 2000) and long-term memory acquisition (Kuhbandner et al. 2016).

EXAMS ALLOW STUDENTS TO DEMONSTRATE PERFORMANCE UNDER PRESSURE

Some argue that examinations reflect real-life situations in the workplace, especially in fields such as the medical and health professions, where information and facts must often be recalled, and decisions made, under time-pressure and without recourse to materials (Van Bergen and Lane 2014; Durning et al. 2016). This is considered an essential skill by employers in a limited set of academic disciplines. However, it is debatable as to whether performance in an examination is a good proxy for performance under pressure in the workplace since the artificially constructed nature of the exam format is unlikely to authentically reflect a genuine workplace situation.

EXAMS GUARD AGAINST CONTRACT CHEATING

One of the most common arguments in favour of high-stakes summative examinations is the belief that they are more effective than other forms of assessment at preventing contract cheating. This is one reason why invigilated closed-book examination formats are favoured: students complete these assessments in a tightly controlled environment, providing photo ID and completing the examination in an open public space under close observation. This ought to minimise opportunities for cheating and plagiarism (Van Bergen and Lane 2014). Nevertheless, it is evident that even these tightly controlled contexts do not provide reliable protection against academic misconduct and cheating (Lancaster and Clarke 2017). Indeed, contract cheating in relation to examinations appears to be prevalent, involving behaviours that range from collusion to impersonation (Bretag, Harper, Burton, Ellis, Newton, Rozenberg, et al. 2019), facilitated by the apparent ease with which university student identification cards can be forged (Potaka and Huang 2015). Sheard & Dick (2003) estimated that the frequency of cheating in examinations in a cohort of graduate students in IT courses approached ten percent, while in McCabe's (2005) study of 64,000 North American university students, over one-third admitted to some form of exam cheating. Bretag et al (2019) found in a large survey of Australian universities that students participated in undetected cheating in invigilated examinations at higher rates than any other type of cheating,

including contract cheating in written assessments. The frequency of academic misconduct in examinations has undoubtedly increased with the move to online learning during the COVID-19 pandemic (Peh, Cerimagic, and Conejos 2021; Reedy et al. 2021; Hill, Mason, and Dunn 2021), and challenges around online proctoring (Raman et al. 2021), mean that there is only minimal capacity to ensure academic integrity in closed-book online examinations.

Biggs (2001) argues that minimisation of the risk of contract cheating is a leading "distorted priority" for the ongoing use of high-stakes examinations. He suggests that "invigilated examinations are hard to justify educationally but are useful logistically and for assuring the public that plagiarism is under control" (2001, 234). Noting that students perceive contract cheating to be most likely for heavily weighted assignments, Bretag et al (2019) similarly suggest that "examinations provide universities and accrediting bodies with a false sense of security" and that "an over reliance on examinations, without a thorough and comprehensive approach to integrity, is likely to lead to more cheating, not less."

EXAMS ARE EFFICIENT TO ADMINISTER AND GRADE

Decreased funding, rising costs and growing student numbers, mean that universities are under growing pressure to prioritise assessment methods that are efficient and cost-effective. A perceived advantage of examinations is their ease of delivery and marking (Ramsden 2003), and there is no doubt that it is less time consuming to mark examinations, especially those that mainly use multiple choice questions, than many other forms of assessment. Nevertheless, considerations of scale and administrative efficiency must be very carefully weighed against the pedagogical merits of such assessments if an institution is genuinely committed to excellent learning experiences for its students.

DRAWBACKS OF HIGH-STAKES SUMMATIVE EXAMS

EXAMS HAVE LIMITED CAPACITY TO ASSESS DEEP LEARNING

One of the main drawbacks of examinations relates to the limited range of skills and knowledge they can develop and assess. Examinations tend to encourage surface-level learning, which occurs when students are directed to memorise and reproduce a series of facts without a more complex understanding of the ways in which ideas and concepts are integrated and make sense in the wider world (Gibbs 1992). Content analyses of examinations in university science and medicine courses, for example, have shown that most questions merely test the isolated recall of factual knowledge (Ramsden 2003). In part, such limitations are an outcome of poorly designed examinations that purely test the recall of information. Examinations can be designed to promote higher levels of thinking, such as by posing questions that demand in-depth analysis and synthesis, and by placing questions within contexts relevant to the field being studied (McConnell, St-Onge, and Young 2015; Villarroel et al. 2020). The format of the examination also determines the range of skills and knowledge that can be fostered and assessed. For example, open book and take-home examinations that allow students to engage with materials and integrate knowledge have a greater potential to measure the application and synthesis of knowledge than closed-book invigilated examinations (Durning et al. 2016), and to assess authentic tasks using real-world scenarios (Deneen 2020). However, even when they are designed to promote higher order thinking, there are limitations as to the kinds of skills examinations can assess, given that their format is almost exclusively written.

EXAMS ARE UNRELIABLE MEASURES OF STUDENT ACHIEVEMENT

Examinations have long been criticised for their low reliability, defined as their ability to produce consistent and dependable results. Low reliability of examination performance can result from a range of factors, related to examinees, examiners, the subject being examined, the test items, and how the examination is scored (Haertel 2006). For instance, the same examinee may perform differently on an examination because of their psychological or physical health, the conditions of the examination space, or their familiarity with the test items selected for the exam. There are also many factors that can affect the performance and judgement of the examiner, including bias, inconsistency, or rater drift (Hartog and Rhodes 1936; Cox 1973; Kellaghan and Greaney 2019; Knight 2002). Finally, the number of test items, whether rubrics are used, and how many items are included in the examination can all affect how it is marked. Not all of these issues are unique to examinations; however, the high-stakes nature of these assessments mean that low reliability is potentially highly problematic.

EXAMS LACK AUTHENTICITY/RELEVANCE AND DO NOT MIRROR 'REAL WORLD' CHALLENGES

A recurring argument against the use of examinations is their lack of authenticity and limited capacity to foster the kinds of skills and knowledge students will need in their future careers, which are more likely to require 'generic' skills such as critical thinking, problem solving and communication, than the ability to recall facts (Boud and Falchikov 2006; Gibbs and Lucas 1997). It is also highly unlikely that summative exams will prepare students to be life-long learners, because they position students as passive recipients of feedback without encouraging them to judge the quality of their work, or to apply the feedback they receive (Boud 2018; Boud and Falchikov 2006). The results of summative examinations are final and preclude any opportunity for students to learn from mistakes or improve their performance (Knight 2002). The capacity to reflect upon, critically appraise and improve ones' own work are likely to be essential to students' future lives and careers. Examination formats are therefore poorly aligned with the imperative for students to develop the capacity to self-assess and to receive and implement frequent, timely feedback.

Moreover, the high-stakes nature of most summative examinations prevents students from adopting an experimental approach to learning, which is a key desired graduate capability. Making mistakes and experiencing misconceptions are an essential part of learning (Metcalfe 2017; Verdake et al. 2017), which is why it is important that sufficient opportunities are provided for students to engage in low-stakes (or no-stakes) assessments early in a subject, and given sufficient opportunity to use 'error detection' (Biggs and Tang 2007) as the basis for correcting and learning from their mistakes.

EXAMS ELEVATE STRESS AND REDUCE WELLBEING

High-stakes examinations are frequently associated with psychological distress and anxiety (Kellaghan and Greaney 2019). This issue has come to the fore in recent years with an increased focus on the role that curriculum and assessment design play in supporting student mental wellbeing (Baik, Larcombe, and Brooker 2019; Slavin, Schindler, and Chibnall 2014). Physiological measures of stress including cardiovascular parameters and stress hormones are higher during examination periods compared to outside these periods (Maes et al. 1998; Wolf and Smith 1995; Weekes et al. 2006; Zhang et al. 2011; Fejes, Ábrahám, and Légrády 2020). Students also self-report higher levels of anxiety during examination periods (Zhang et al. 2011; Ballen, Salehi, and Cotner 2017); a reliable correlate of physiological stress (Roos et al. 2021).

The negative impact of high-stakes examinations on student health and mental wellbeing is concerning in and of itself, but the stress and anxiety caused by examinations also affects their reliability as a measure of student learning. Higher levels of anxiety are linked to reduced concentration (Fernández-Castillo and Caurcel 2015) and performance (Wolf and Smith 1995) during examinations, and there is evidence to suggest that stress impairs memory retrieval (Vogel and Schwabe 2016). Students more prone to examination anxiety are also more likely to have lower self-esteem and sleep less during examination periods, which contributes further to reduced wellbeing and impedes examination performance (Fernández-Castillo and Caurcel 2019). It is likely that higher-weighted examinations have stronger negative impacts on students' wellbeing, due to the increased perception of consequences of the outcomes (Franke 2018; Salehi et al. 2019; Wolf and Smith 1995). While some argue that examination anxiety is useful as it promotes study and preparation (Hamzah et al. 2018), and can increase performance (Shean 2019), such potentially positive gains need to be weighed carefully against the negative impacts on mental and physical health. There are also a range of cultural and genetic factors that exacerbate experiences of examination stress (Zhang et al. 2011), meaning that students will be affected unequally.

EXAMS PROMOTE ACADEMIC INEQUITY

It is known that students perform differently under time pressure (De Paola and Gioia 2016), and there is also considerable evidence to suggest that the examination method has equity implications, which intersects with impacts on wellbeing and student learning. Studies within the STEM disciplines suggest that examinations differentially affect students based on their gender, finding that women tend to experience higher levels of assessment anxiety leading to lower wellbeing and reduced concentration during an examination, and resulting in lower performance (Fernández-Castillo and Caurcel 2015; 2019; Roos et al. 2021; Salehi et al. 2019), an effect that may be stronger at introductory levels of university (Ballen, Salehi, and Cotner 2017; Salehi et al. 2019). A study by Ballen, Salehi and Cotner (2017) found that women in an introductory biology course underperformed on examinations compared to their male counterparts but outperformed them on combined non-examination methods of assessment, suggesting that there may be correlations between assessment modes and gendered styles of learning.

In addition to potential gender biases, examinations may be biased towards Western students, with equity implications for international students from non-Western countries and for Indigenous students. Richardson (2015a; 2015b) suggests it is likely that the under-attainment of students from ethnic minorities is connected to assessment methods, while many have argued that examinations disadvantage Indigenous students (Preston and Claypool 2021; Claypool and Preston 2013; Klenowski 2009; Trumbull and Nelson-Barber 2019), as they tend to promote Western intellectual knowledge and values by supporting the view that knowledge can be given, accumulated, and tested in a rational and linear manner. Recent scholarship on inclusive assessment design further argues that examinations fail to meet the needs of student diversity, especially with respect to students with disabilities (Nieminen 2022; Tai et al. 2022; Nieminen and Tuohilampi 2020).

RECONSIDERING THE ROLE OF EXAMINATIONS IN THE CURRICULUM

Our review of the literature suggests that the use of high-stakes examinations may be based on considerations that are not always grounded in empirical evidence, and therefore that the use of such assessments should be carefully reviewed. Such reviews could result in a recommendation to *retain, redesign, reweight or replace* the high-stakes examination as an assessment item.

RETAINING OR REDESIGNING EXAMINATIONS

When well-designed and used within a balanced and diversified assessment regime, examinations have an ongoing role to play in the curriculum. Retaining examinations may be justified in many subjects. However, where examinations are retained in the curriculum, it is imperative that they are well-designed, for both pedagogical and security reasons. In some cases, it may be useful to consider redesigning examinations to reduce the emphasis on memory recall and increase opportunities for students to demonstrate higher-order thinking skills such as analysis and synthesis. For example, research shows that the use of short-answer questions as well as context-rich multiple-choice questions that require the application of knowledge enhance learning relative to multiple-choice questions that require the recall of facts (McConnell, St-Onge, and Young 2015). Improving the authenticity of examinations by designing examination questions that reflect real-life situations in the workplace and that require evaluative judgement have also been shown to support deep approaches to learning (Villarroel et al. 2020).

Effective design is also important to avoid obvious mistakes that make examinations vulnerable to student 'test-wiseness' (Towns and Robinson 1993) or weaken their security (Dawson 2020). For example, questions that assess memory recall should be replaced/balanced with questions that require application, analysis, and synthesis. Research has shown that the benefits of test-enhanced learning are greater when questions are context-rich and require the application of knowledge, than when they simply require the recognition of facts (McConnell, St-Onge, and Young 2015), and that short answer questions tend to be more beneficial than multiple choice questions (McDaniel et al. 2007). Well-designed open-book examinations may offer an advance on the traditional closed-book examination and are known to have a range of educational benefits including a greater tendency to engage higher-order skills over rote learning (Deneen 2020). There are also a range of purposes for which examinations might be employed other than to test students in high-weighted summative assessments. For example, shorter examinations or tests scheduled throughout the semester can be useful to build student learning over time, while groupwork examinations can be employed to engage students in collaborative learning tasks. To minimize opportunities for contract cheating, examination papers should not be reused, online tests should not be unsupervised, and low-level or 'one right answer' tasks and questions should be avoided (Dawson 2020).

REWEIGHTING AND REPLACING EXAMINATIONS

By reducing the emphasis and weighting of high-stakes summative examinations in the curriculum there is an opportunity to employ a range of creative approaches to improve assessment and feedback practices at the University. This may be achieved by reweighting examinations to remove their high-stakes consequences and make room for a broader range of assessment methods, or in some cases by replacing them with other assessment options. Below we outline some alternatives to high-stakes summative examinations.

Diversified assessment formats

It is crucial that assessments allow students to develop and demonstrate their skills and knowledge using a variety of different formats. While summative examinations are often constrained by their predominantly written format, assessments scheduled throughout the semester can take more diverse forms. There are more opportunities for group activities and collaboration between students, which fosters skills such as communication and teamwork, and can increase student confidence and engagement (Crooks 1988). Formative or continuous coursework assessments also have the potential to be more creative, to require students to apply their skills in more varied ways, and to be rewarded for originality rather than conformity (Ramsden 2003). Assessment tasks that engage students in self-assessment, peer-peer learning, problem-based learning, and authentic learning all foster important graduate capabilities, and can be developed through a broad range of assessable activities. These include in-class and online tests and quizzes, online social annotations and discussions, student presentations, debates, simulations, clinical exercises, portfolios, critical reflections, research projects, and inquiry-based group projects.

Enhancing opportunities for formative assessment and feedback

The capacity for formative assessments to better allow students to demonstrate their ability is evidenced by the numerous studies showing that student performance is improved in assessments that take place outside of the examination context (Bridges et al. 2002; Richardson 2015a; Yorke, Bridges, and Woolf 2000; Simonite 2003). Yorke et al (2000) argue that this is because students are unable to perform optimally under the time-pressure and artificial conditions of examinations, which prevent students from enhancing their performance by accessing information, taking time to develop their work, or learn through collaborative working. For Gibbs and Lucas (1997), the increase in academic attainment in continuous coursework can be attributed to the capacity for students to obtain and implement feedback. They argue that the grades for coursework are more reliable as a measure of attainment and that they correlate better with long-term measures of learning. Similarly, Bridges et al (2002) suggest that students achieve better outcomes in coursework assignments because they respond positively to having more control over their own work, and because they have access to both resources and help, although they acknowledge such help may be in the form of authorised guidance, or unauthorised collusion or plagiarism.

Discouraging and minimising opportunities for cheating

Concerns about academic misconduct can potentially be alleviated by assessment designs that minimize opportunities for cheating (Dawson 2020). While the increasing prevalence of "assignment outsourcing" by ghost writers, essay mills, and AI platforms such as ChatGPT is well known (Ali and Alhassan 2021; Awdry 2021; Huang 2023), careful assessment design may help combat or reduce opportunities and incentives for cheating in a range of ways (Baird and Clare 2017). For instance, assessment tasks or questions that ask students to reflect or draw on personal circumstances or experiences (Sutherland-Smith 2008), local contexts or environments, or assessment tasks that are conducted within a specific class or tutorial activity should generally be more difficult to procure from external sources than standard essays on common topics. Similarly, where tasks involve repeated contributions (reflective journals, blogs), audit trails of progress, or other forms of 'authentic' assessment, they ought to be difficult or costly to obtain from external providers. Finally, authorship of some assessment tasks such as vivas, individual or group oral presentations, or video presentations can be verified with a relatively high degree of confidence. It is nevertheless important to recognise that ultimately, no assessment type, whether high-stakes examination or 'authentic' task, is immune to outsourcing (Bretag, Harper, Burton, Ellis, Newton, van Haeringen, et al. 2019; Ellis et al. 2020). Thus, academic integrity considerations are ultimately not compelling reasons either for retaining or abandoning high-stakes examinations.

Scalable and efficient assessment methods

Can assessment alternatives to high-stakes examinations provide similar scalability and administrative convenience? There is no doubt that some of the assessment formats that are least likely to be outsourced (e.g. vivas) will also be most resource-intensive to administer and grade. Not surprisingly, educators' willingness to consider such alternatives correlates strongly with perceptions of institutional support (Bretag, Harper, Burton, Ellis, Newton, van Haeringen, et al. 2019). However, there are promising examples of educational technology that can assist with the administrative burden of distributing and grading assessments other than high-stakes examinations at scale. Examples include platforms that support peer assessment (Søndergaard and Mulder 2012), social annotation (Miller et al. 2018), personalising feedback through the use of digital recordings (Ryan, French, and Kennedy 2021) and automated feedback and grading (Cavalcanti et al. 2021; Hegarty-Kelly and Mooney 2021; Kumar and Boulanger 2021).

Programmatic assessment

Programmatic assessment also offers an emerging approach to assessment design that has the potential to both increase assessment security (Dawson 2020) and to reduce the reliance on high-stakes summative examinations by diversifying assessment methods across the curriculum of an entire program (Baartman, van Schilt-Mol, and van der Vleuten 2022; Heeneman et al. 2021). Programmatic assessment refers to a conceptualisation of assessment as a continuous and intentional program across a degree, rather than as a series of individual tasks or courses (Schuwirth and Van der Vleuten 2011). In contrast to traditional assessment regimes that tend to comprise low stakes formative assessments and high-stakes summative examinations within single units or subjects, thinking about assessment at the program level avoids binary classification and allows for a continuum from low to high-stakes depending on learner progress (Heeneman et al. 2021). While not trivial to implement, a programmatic approach allows for intentional emphasis on low-weighted assessments in the foundational years of a degree, placing more emphasis on assessment tasks that foster the development of student capabilities and cohort connections. Examinations can then be employed at key moments for accreditation purposes (although other methods might also fulfil this role), but final summative examinations would no longer be the default assessment mode.

CONCLUSION

When they are well designed and used within a balanced and diverse program of assessment, examinations have an ongoing role to play in university curricula in some disciplines. However, the empirical evidence suggests that an over-use of examinations is problematic for a range of reasons that include their encouragement of surface learning, inauthentic relationship to the workplace, lack of alignment with desired graduate outcomes, deterrence of experimental approaches to learning, negative impact on student wellbeing and equity implications. By contrast, it is difficult to find either evidence or opinion to support the use of high-stakes examinations in the educational literature, which suggests that they are predominantly used for pragmatic reasons rather than pedagogical ones.

The use of examinations becomes particularly problematic when they dominate the curriculum at the expense of other valuable forms of assessment and deprive students of opportunities to engage in assessment tasks that develop a broader range of skills that align with desired graduate outcomes and promote a culture of life-long learning. It is critical that students are provided with ample opportunities to develop their learning and build diverse skills by engaging with a sufficient variety of assessment methods, which is often precluded by the dominance of high-stakes examinations in the curriculum. In redesigning the curriculum to achieve a greater variety in assessment methods therefore, the weighting given to summative examinations is an important consideration. It has been suggested that to discourage ‘cramming’, ‘the final should count for no more than one-third of the course grade’ (Davis 1993, 288), yet it has also been found that an overuse of formative assessment tasks can increase stress and reduce enjoyment of the university experience (Harland et al. 2015). It can therefore be a ‘balancing act’ for educators (Franke 2018) to determine a fair weighting that simultaneously limits the stress attached to high-stakes summative assessments *and* ensures that there is sufficient weight to motivate students to study and allow them to demonstrate their achievements.

REFERENCES

Ali, Holi Ibrahim Holi, and Awad Alhassan. 2021. “Fighting Contract Cheating and Ghostwriting in Higher Education: Moving towards a Multidimensional Approach.” Edited by Sammy King Fai HUI. *Cogent Education* 8 (1): 1885837. https://doi.org/10.1080/2331186X.2021.1885837.

Awdry, Rebecca. 2021. “Assignment Outsourcing: Moving beyond Contract Cheating.” *Assessment & Evaluation in Higher Education* 46 (2): 220–35. https://doi.org/10.1080/02602938.2020.1765311.

Baartman, Liesbeth, Tamara van Schilt-Mol, and Cees van der Vleuten. 2022. “Programmatic Assessment Design Choices in Nine Programs in Higher Education.” *Frontiers in Education* 7 (October): 931980. https://doi.org/10.3389/feduc.2022.931980.

Baik, Chi, Wendy Larcombe, and Abi Brooker. 2019. “How Universities Can Enhance Student Mental Wellbeing: The Student Perspective.” *Higher Education Research & Development* 38 (4): 674–87. https://doi.org/10.1080/07294360.2019.1576596.

Baird, Michael, and Joseph Clare. 2017. “Removing the Opportunity for Contract Cheating in Business Capstones: A Crime Prevention Case Study.” *International Journal for Educational Integrity* 13 (1): 6. https://doi.org/10.1007/s40979-017-0018-1.

Ballen, Cissy J., Shima Salehi, and Sehoya Cotner. 2017. “Exams Disadvantage Women in Introductory Biology.” *PLoS ONE* 12 (10). https://doi.org/10.1371/journal.pone.0186419.

Bassey, Michael. 1971. “The Assessments of Students by Formal Assignments.” New Zealand University Students Association. http://www.tandfonline.com/doi/abs/10.1080/02602938.2014.919628.

Biggs, John. 2001. “The Reflective Institution: Assuring and Enhancing the Quality of Teaching and Learning 1.” Vol. 41.

———. 2003. *Teaching for Quality Learning at University: What the Student Does*. 2nd ed. Buckingham; Philadelphia, PA: Society for Research into Higher Education: Open University Press.

Biggs, John, and Catherine Tang. 2007. “Aligning Assessment Tasks with Intended Learning Outcomes: Principles.” In *Teaching for Quality Learning at University: What the Student Does*, 3rd ed.

Black et al, ed. 2009. *Assessment for Learning: Putting It into Practice*. Reprinted. Maidenhead: Open Univ. Press.

Boud, David. 2000. “Sustainable Assessment: Rethinking Assessment for the Learning Society.” *Studies in Continuing Education* 22 (2): 151–67. https://doi.org/10.1080/713695728.

———. 2013. *Enhancing Learning Through Self-Assessment*. 0 ed. Routledge. https://doi.org/10.4324/9781315041520.

———, ed. 2018. *Developing Evaluative Judgement in Higher Education*. Abingdon, Oxon ; New York, NY: Routledge.

Boud, David, and Nancy Falchikov. 2006. “Aligning Assessment with Long-Term Learning.” *Assessment and Evaluation in Higher Education* 31 (4): 399–413. https://doi.org/10.1080/02602930600679050.

Boud, David, and Rebeca Soler. 2016. “Sustainable Assessment Revisited.” *Assessment & Evaluation in Higher Education* 41 (3): 400–413. https://doi.org/10.1080/02602938.2015.1018133.

Bretag, Tracey, Rowena Harper, Michael Burton, Cath Ellis, Philip Newton, Karen van Haeringen, Sonia Saddiqui, and Pearl Rozenberg. 2019. “Contract Cheating and Assessment Design: Exploring the Relationship.” *Assessment & Evaluation in Higher Education* 44 (5): 676–91. https://doi.org/10.1080/02602938.2018.1527892.

Bretag, Tracey, Rowena Harper, Michael Burton, Cath Ellis, Philip Newton, Pearl Rozenberg, Sonia Saddiqui, and Karen van Haeringen. 2019. “Contract Cheating: A Survey of Australian University Students.” *Studies in Higher Education* 44 (11): 1837–56. https://doi.org/10.1080/03075079.2018.1462788.

Bridges, Paul, Angela Cooper, Peter Evanson, Chris Haines, Don Jenkins, David Scurry, Harvey Woolf, and Mantz Yorke. 2002. “Coursework Marks High, Examination Marks Low: Discuss.” *Assessment & Evaluation in Higher Education* 27 (1): 35–48. https://doi.org/10.1080/02602930120105045.

Butler, Andrew C., and Henry L. Roediger. 2007. “Testing Improves Long-Term Retention in a Simulated Classroom Setting.” *European Journal of Cognitive Psychology* 19 (4–5): 514–27. https://doi.org/10.1080/09541440701326097.

Carless, David, Susan M. Bridges, Cecilia Ka Yuk Chan, and Rick Glofcheski, eds. 2017. *Scaling up Assessment for Learning in Higher Education*. Vol. 5. The Enabling Power of Assessment. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-10-3045-1.

Cavalcanti, Anderson Pinheiro, Arthur Barbosa, Ruan Carvalho, Fred Freitas, Yi-Shan Tsai, Dragan Gašević, and Rafael Ferreira Mello. 2021. "Automatic Feedback in Online Learning Environments: A Systematic Literature Review." *Computers and Education: Artificial Intelligence* 2: 100027. https://doi.org/10.1016/j.caeai.2021.100027.

Claypool, Tim R., and Jane P. Preston. 2013. "Redefining Learning and Assessment Practices Impacting Aboriginal Students: Considering Aboriginal Priorities Via Aboriginal and Western Worldviews." *In Education* 17 (3). https://doi.org/10.37119/ojs2011.v17i3.74.

Cox, R.J. 1973. "Traditional Examinations in a Changing Society." *Universities Quarterly* Spring: 200–216.

Crooks, Terence J. 1988. "The Impact of Classroom Evaluation Practices on Students." Vol. 58. Source: Review of Educational Research. Winter.

Davis, Barbara Gross. 1993. *Tools for Teaching*. The Jossey-Bass Higher and Adult Education Series. San Francisco, CA: Jossey-Bass.

Dawson, Phillip. 2020. "Structuring and Designing Assessment for Security." In *Defending Assessment Security in a Digital World: Preventing E-Cheating and Supporting Academic Integrity in Higher Education*, 1st ed. Abingdon, Oxon ; New York, NY : Routledge, 2021.: Routledge. https://doi.org/10.4324/9780429324178.

De Paola, Maria, and Francesca Gioia. 2016. "Who Performs Better under Time Pressure? Results from a Field Experiment." *Journal of Economic Psychology* 53 (April): 37–53. https://doi.org/10.1016/j.joep.2015.12.002.

Deneen, Chris. 2020. "Assessment Considerations in Moving from Closed-Book to Open-Book Exams." Melbourne CSHE. https://melbourne-cshe.unimelb.edu.au/__data/assets/pdf_file/0010/3341944/closed-book-to-open-book-exam_final.pdf.

Durning, Steven J., Ting Dong, Temple Ratcliffe, Lambert Schuwirth, Anthony R. Artino, John R. Boulet, and Kevin Eva. 2016. "Comparing Open-Book and Closed-Book Examinations: A Systematic Review." *Academic Medicine* 91 (4): 583–99. https://doi.org/10.1097/ACM.0000000000000977.

Ecclestone, Kathryn. 1999. "Empowerng or Ensnaring?: The Implications of Outcome-based Assessment in Higher Education." *Higher Education Quarterly* 53 (1): 29–48. https://doi.org/10.1111/1468-2273.00111.

Ellis, Cath, Karen van Haeringen, Rowena Harper, Tracey Bretag, Ian Zucker, Scott McBride, Pearl Rozenberg, Phil Newton, and Sonia Saddiqui. 2020. "Does Authentic Assessment Assure Academic Integrity? Evidence from Contract Cheating Data." *Higher Education Research & Development* 39 (3): 454–69. https://doi.org/10.1080/07294360.2019.1680956.

Entwistle, N. J., and Abigail Entwistle. 1991. "Contrasting Forms of Understanding for Degree Examinations: The Student Experience and Its Implications." *Higher Education* 22 (3): 205–27. https://doi.org/10.1007/BF00132288.

Fejes, Imola, György Ábrahám, and Péter Légrády. 2020. "The Effect of an Exam Period as a Stress Situation on Baroreflex Sensitivity among Healthy University Students." *Blood Pressure* 29 (3): 175–81. https://doi.org/10.1080/08037051.2019.1710108.

Fernández-Castillo, Antonio, and María J. Caurcel. 2015. "State Test-Anxiety, Selective Attention and Concentration in University Students." *International Journal of Psychology* 50 (4): 265–71. https://doi.org/10.1002/ijop.12092.

———. 2019. "Self-Esteem, Hours of Sleep and State-Anxiety before Academic Tests." *Revista Argentina de Clinica Psicologica* 28 (4): 348–55. https://doi.org/10.24205/03276716.2017.1039.

Fook, C.Y., and G.K. Sidhu. 2014. "Assessment Practices in Higher Education in United States." *Procedia - Social and Behavoiral Sceinces* 123: 299–306.

Franke, Matthew. 2018. "Final Exam Weighting as Part of Course Design." *Teaching and Learning Inquiry* 6 (1): 91–103. https://doi.org/10.20343/teachlearninqu.6.1.9.

Gibbs, Graham. 1992. *Improving the Quality of Student Learning*. Bristol: Teachnical and Educational Services Ltd.

Gibbs, Graham, and Lisa Lucas. 1997. “Coursework Assessment, Class Size and Student Performance: 1984-94.” *Journal of Further and Higher Education* 21 (2): 183–92. https://doi.org/10.1080/0309877970210204.

Greene, Edward B. 1931. “The Retention of Information Learned in College Courses.” *The Journal of Educational Research* 24 (4): 262–73. https://doi.org/10.1080/00220671.1931.10880208.

Haertel, E. 2006. “Reliability.” In *Educational Measurement, 4th Ed., Edited by R. L. Brennan, Pp65–100.* Westport, CT: American Council on Education and Praeger.

Hamzah, Farrahdilla, Khairi Che Mat, Vidya Bhagat, and Nursyakilah Mahyiddin. 2018. “Test Anxiety and Its Impact on First Year University Students and the over View of Mind and Body Intervention to Enhance Coping Skills in Facing Exams.” *Research Journal of Pharmacy and Technology* 11 (6): 2220–28. https://doi.org/10.5958/0974-360X.2018.00411.0.

Harland, Tony, Angela McLean, Rob Wass, Ellen Miller, and Kwong Nui Sim. 2015. “An Assessment Arms Race and Its Fallout: High-Stakes Grading and the Case for Slow Scholarship.” *Assessment and Evaluation in Higher Education* 40 (4): 528–41. https://doi.org/10.1080/02602938.2014.931927.

Hartog, P, and E.C. Rhodes. 1936. *An Examination of Examinations*. London: Macmmillan.

Hartwig, Marissa K., and John Dunlosky. 2012. “Study Strategies of College Students: Are Self-Testing and Scheduling Related to Achievement?” *Psychonomic Bulletin & Review* 19 (1): 126–34. https://doi.org/10.3758/s13423-011-0181-y.

Heeneman, Sylvia, Lubberta H. de Jong, Luke J. Dawson, Tim J. Wilkinson, Anna Ryan, Glendon R. Tait, Neil Rice, Dario Torre, Adrian Freeman, and Cees P. M. van der Vleuten. 2021. “Ottawa 2020 Consensus Statement for Programmatic Assessment – 1. Agreement on the Principles.” *Medical Teacher* 43 (10): 1139–48. https://doi.org/10.1080/0142159X.2021.1957088.

Hegarty-Kelly, Emlyn, and Dr Aidan Mooney. 2021. “Analysis of an Automatic Grading System within First Year Computer Science Programming Modules.” In *Computing Education Practice 2021*, 17–20. Durham United Kingdom: ACM. https://doi.org/10.1145/3437914.3437973.

Henderson, Michael, Rola Ajjawi, David Boud, and Elizabeth Molloy, eds. 2020. *The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners*. Cham, Switzerland: Palgrave Macmillan.

Hill, Guzyal, Jon Mason, and Alex Dunn. 2021. “Contract Cheating: An Increasing Challenge for Global Academic Community Arising from COVID-19.” *Research and Practice in Technology Enhanced Learning* 16 (1): 24. https://doi.org/10.1186/s41039-021-00166-8.

Huang, K. 2023. "Alarmed by A.I. Chatbots, Universities Start Revamping How They Teach". *New York Times. https://www.nytimes.com/2023/01/16/technology/chatgpt-artificial-intelligence-universities.html*

Jones, Emma, Michael Priestley, Liz Brewster, Susan J. Wilbraham, Gareth Hughes, and Leigh Spanner. 2021. “Student Wellbeing and Assessment in Higher Education: The Balancing Act.” *Assessment and Evaluation in Higher Education* 46 (3): 438–50. https://doi.org/10.1080/02602938.2020.1782344.

Jones, Harriet, Beth Black, Jon Green, Phil Langton, Stephen Rutherford, Jon Scott, and Sally Brown. 2015. “Indications of Knowledge Retention in the Transition to Higher Education.” *Journal of Biological Education* 49 (3): 261–73. https://doi.org/10.1080/00219266.2014.926960.

Karpicke, Jeffrey D., and Henry L. Roediger. 2008. “The Critical Importance of Retrieval for Learning.” *Science* 319 (5865): 966–68. https://doi.org/10.1126/science.1152408.

Kellaghan, Thomas, and Vincent Greaney. 2019. *Public Examinations Examined*. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-1418-1.

Kickert, R., M. Meeuwisse, K. M. Stegers-Jager, P. Prinzie, and L. R. Arends. 2022. “Curricular Fit Perspective on Motivation in Higher Education.” *Higher Education* 83 (4): 729–45. https://doi.org/10.1007/s10734-021-00699-3.

Klenowski, Val. 2009. “Australian Indigenous Students: Addressing Equity Issues in Assessment.” *Teaching Education* 20 (1): 77–93.

Knight, Peter T. 2002. “Summative Assessment in Higher Education: Practices in Disarray.” *Studies in Higher Education* 27 (3): 275–86. https://doi.org/10.1080/03075070220000662.

Kuhbandner, Christof, Alp Aslan, Kathrin Emmerdinger, and Kou Murayama. 2016. "Providing Extrinsic Reward for Test Performance Undermines Long-Term Memory Acquisition." *Frontiers in Psychology* 7 (February). https://doi.org/10.3389/fpsyg.2016.00079.

Kumar, Vivekanandan S., and David Boulanger. 2021. "Automated Essay Scoring and the Deep Learning Black Box: How Are Rubric Scores Determined?" *International Journal of Artificial Intelligence in Education* 31 (3): 538–84. https://doi.org/10.1007/s40593-020-00211-5.

Lancaster, Thomas, and Robert Clarke. 2017. "Rethinking Assessment by Examination in the Age of Contract Cheating." In *Plagiarism across Europe and Beyond 2017—Conference Proceedings*, 215–28.

Maes, Michael, Marc Van Der Planken, Ann Van Gastel, Karen Bruyland, Fran Van Hunsel, Hugo Neels, Dirk Hendriks, et al. 1998. "Influence of Academic Examination Stress on Hematological Measurements in Subjectively Healthy Volunteers." *Psychiatry Research* 80 (3): 201–12. https://doi.org/10.1016/S0165-1781(98)00059-6.

Martin, Linley, and Marian Mahat. 2017. "The Assessment of Learning Outcomes in Australia: Finding the Holy Grail." *AERA Open* 3 (1): 233285841668890. https://doi.org/10.1177/2332858416688904.

Marton, F, and R Säljö. 1997. "Approaches to Learning." In *The Experience of Learning. Implications for Teaching and Studying in Higher Education.*, edited by F Marton, D Hounsell, and N Entwistle, 39–59. Edinburgh: Scottish Academic Press. http://www.docs.hss.ed.ac.uk/iad/Learning_teaching/Academic_teaching/Resources/Experience_of_learning/EoLChapter3.pdf.

McCabe, Donald L. 2005. "Cheating among College and University Students: A North American Perspective." *International Journal for Educational Integrity* 1 (1). https://doi.org/10.21913/IJEI.v1i1.14.

McConnell, M M, C St-Onge, and M E Young. 2015. "The Benefits of Testing for Learning on Later Performance." *ADVANCES IN HEALTH SCIENCES EDUCATION* 20 (2): 305–20. https://doi.org/10.1007/s10459-014-9529-1.

McDaniel, Mark A., Janis L. Anderson, Mary H. Derbish, and Nova Morrisette. 2007. "Testing the Testing Effect in the Classroom." *European Journal of Cognitive Psychology* 19 (4–5): 494–513. https://doi.org/10.1080/09541440701326154.

Metcalfe, Janet. 2017. "Learning from Errors." *Annual Review of Psychology* 68 (1): 465–89. https://doi.org/10.1146/annurev-psych-010416-044022.

Miller, Kelly, Brian Lukoff, Gary King, and Eric Mazur. 2018. "Use of a Social Annotation Platform for Pre-Class Reading Assignments in a Flipped Introductory Physics Class." *Frontiers in Education* 3 (March): 8. https://doi.org/10.3389/feduc.2018.00008.

Morris, Rebecca, Thomas Perry, and Lindsey Wardle. 2021. "Formative Assessment and Feedback for Learning in Higher Education: A Systematic Review." *Review of Education* 9 (3). https://doi.org/10.1002/rev3.3292.

Nieminen, Juuso Henrik. 2022. "Assessment for Inclusion: Rethinking Inclusive Assessment in Higher Education." *Teaching in Higher Education*, January, 1–19. https://doi.org/10.1080/13562517.2021.2021395.

Nieminen, Juuso Henrik, and Laura Tuohilampi. 2020. "'Finally Studying for Myself' – Examining Student Agency in Summative and Formative Self-Assessment Models." *Assessment & Evaluation in Higher Education* 45 (7): 1031–45. https://doi.org/10.1080/02602938.2020.1720595.

Pascoe, Michaela C., Sarah E. Hetrick, and Alexandra G. Parker. 2020. "The Impact of Stress on Students in Secondary School and Higher Education." *International Journal of Adolescence and Youth* 25 (1): 104–12. https://doi.org/10.1080/02673843.2019.1596823.

Peh, Luke Lu Chang, Sabina Cerimagic, and Sheila Conejos. 2021. "Challenges of Running Online Exams and Preventing Academic Dishonesty during the Covid-19 Pandemic." *Journal of Learning Development in Higher Education*, no. 22.

Potaka, Elisa, and Cecily Huang. 2015. "Pens For Hire: How Students Cheat, and How They Get Away with It." *SBS*, November 15, 2015. https://www.sbs.com.au/news/the-feed/article/pens-for-hire-how-students-cheat-and-how-they-get-away-with-it/5v3erlpij.

Preston, Jane P., and Tim R. Claypool. 2021. "Analyzing Assessment Practices for Indigenous Students." *Frontiers in Education* 6 (July). https://doi.org/10.3389/feduc.2021.679972.

Raman, Raghu, Sairam B, Veena G, Hardik Vachharajani, and Prema Nedungadi. 2021. “Adoption of Online Proctored Examinations by University Students during COVID-19: Innovation Diffusion Study.” *Education and Information Technologies* 26 (6): 7339–58. https://doi.org/10.1007/s10639-021-10581-5.

Ramsden, P. 1997. “The Context of Learning in Academic Departments.” In *The Experience of Learning. Implications for Teaching and Studying in Higher Education*, edited by F Marton, D Hounsell, and N Entwistle, 198–217. Edinburgh: Scottish Academic Press. http://www.docs.hss.ed.ac.uk/iad/Learning_teaching/Academic_teaching/Resources/Experience_of_learning/EoLChapter13.pdf.

———. 2003. “Assessing for Understanding.” In *Learning to Teach in Higher Education*. http://ebookcentral.proquest.com','_blank'.

Rawlusyk, P.E. 2018. “Assessment in Higher Education and Student Learning.” *Journal of Instructional Pedagogies* 21.

Rawson, Katherine A., John Dunlosky, and Sharon M. Sciartelli. 2013. “The Power of Successive Relearning: Improving Performance on Course Exams and Long-Term Retention.” *Educational Psychology Review* 25 (4): 523–48. https://doi.org/10.1007/s10648-013-9240-4.

Reedy, Alison, Darius Pfitzner, Laura Rook, and Leonie Ellis. 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). https://doi.org/10.1007/s40979-021-00075-9.

Richardson, John T.E. 2015a. “Coursework versus Examinations in End-of-Module Assessment: A Literature Review.” *Assessment and Evaluation in Higher Education* 40 (3): 439–55. https://doi.org/10.1080/02602938.2014.919628.

———. 2015b. “The Under-Attainment of Ethnic Minority Students in UK Higher Education: What We Know and What We Don’t Know.” *Journal of Further and Higher Education* 39 (2): 278–91. https://doi.org/10.1080/0309877X.2013.858680.

Roediger, Henry L., and Andrew C. Butler. 2011. “The Critical Role of Retrieval Practice in Long-Term Retention.” *Trends in Cognitive Sciences* 15 (1): 20–27. https://doi.org/10.1016/j.tics.2010.09.003.

Roediger, Henry L., and Jeffrey D. Karpicke. 2006. “Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention.” *Psychological Science* 17 (3): 249–55. https://doi.org/10.1111/j.1467-9280.2006.01693.x.

Roos, Anna Lena, Thomas Goetz, Martin Voracek, Maike Krannich, Madeleine Bieg, Amanda Jarrell, and Reinhard Pekrun. 2021. “Test Anxiety and Physiological Arousal: A Systematic Review and Meta-Analysis.” *Educational Psychology Review* 33 (2): 579–618. https://doi.org/10.1007/s10648-020-09543-z.

Ryan, Richard M., and Edward L. Deci. 2000. “Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions.” *Contemporary Educational Psychology* 25 (1): 54–67. https://doi.org/10.1006/ceps.1999.1020.

Ryan, Tracii. 2020. “Effective Feedback in Digital Learning Environments.” Melbourne CSHE. https://melbourne-cshe.unimelb.edu.au/__data/assets/pdf_file/0004/3417079/effective-feedback-in-digital-learning-environments_final.pdf.

Ryan, Tracii, Sarah French, and Gregor Kennedy. 2021. “Beyond the Iron Triangle: Improving the Quality of Teaching and Learning at Scale.” *Studies in Higher Education* 46 (7): 1383–94. https://doi.org/10.1080/03075079.2019.1679763.

Salehi, S, S Cotner, S M Azarin, E E Carlson, M Driessen, V E Ferry, W Harcombe, et al. 2019. “Gender Performance Gaps Across Different Assessment Methods and the Underlying Mechanisms: The Case of Incoming Preparation and Test Anxiety.” *Frontiers in Education* 4. https://doi.org/10.3389/feduc.2019.00107.

Santovena-Casal, S. 2019. “Effects of Continuous Assessment on the Academic Performance of Future Teachers.” *Croatian Journal of Education* 21 (3): 777–822. https://doi.org/10.15516/cje.v21i3.3013.

Schuwirth, Lambert W. T., and Cees P. M. Van der Vleuten. 2011. “Programmatic Assessment: From Assessment of Learning to Assessment for Learning.” *Medical Teacher* 33 (6): 478–85. https://doi.org/10.3109/0142159X.2011.565828.

Shean, Mandie. 2019. “Don’t Calm down! Exam Stress May Not Be Fun but It Can Help You Get Better Marks.” *The Conversation*, October 16, 2019. https://theconversation.com/dont-calm-down-exam-stress-may-not-be-fun-but-it-can-help-you-get-better-marks-124517.

Sheard, Judy, and Martin Dick. 2003. "Influences on Cheating Practice of Graduate Students in It Courses: What Are the Factors?" *Proceedings of the Annual SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiSCE)* 8 (September): 45–49. https://doi.org/10.1145/961290.961527.

Simonite, Vanessa. 2003. "The Impact of Coursework on Degree Classifications and the Performance of Individual Students." *Assessment & Evaluation in Higher Education* 28 (5): 459–70. https://doi.org/10.1080/02602930301675.

Slavin, Stuart J., Debra L. Schindler, and John T. Chibnall. 2014. "Medical Student Mental Health 3.0: Improving Student Wellness Through Curricular Changes." *Academic Medicine* 89 (4): 573–77. https://doi.org/10.1097/ACM.0000000000000166.

Søndergaard, Harald, and Raoul A. Mulder. 2012. "Collaborative Learning through Formative Peer Review: Pedagogy, Programs and Potential." *Computer Science Education* 22 (4): 343–67. https://doi.org/10.1080/08993408.2012.728041.

Sutherland-Smith, W. 2008. *Plagiarism, the Internet, and Student Learning: Improving Academic Integrity.* New York: Routledge.

Tai, Joanna, Rola Ajjawi, Margaret Bearman, David Boud, Phillip Dawson, and Trina Jorre de St Jorre. 2022. "Assessment for Inclusion: Rethinking Contemporary Strategies in Assessment Design." *Higher Education Research & Development*, June, 1–15. https://doi.org/10.1080/07294360.2022.2057451.

Taras, Maddalena. 2010. "Assessment for Learning: Assessing the Theory and Evidence." *Procedia - Social and Behavioral Sciences* 2 (2): 3015–22. https://doi.org/10.1016/j.sbspro.2010.03.457.

Towns, Marcy Hamby, and William R. Robinson. 1993. "Student Use of Test-Wiseness Strategies in Solving Multiple-Choice Chemistry Examinations." *Journal of Research in Science Teaching* 30 (7): 709–22. https://doi.org/10.1002/tea.3660300709.

Trumbull, Elise, and Sharon Nelson-Barber. 2019. "The Ongoing Quest for Culturally-Responsive Assessment for Indigenous Students in the U.S." *Frontiers in Education* 4 (June). https://doi.org/10.3389/feduc.2019.00040.

Van Bergen, P., and R. Lane. 2014. "Exams Might Be Stressful, but They Improve Learning." *The Conversation*, December 19, 2014. https://theconversation.com/exams-might-be-stressful-but-they-improve-learning-35614.

Vaughan, Norman. 2014. "Student Engagement and Blended Learning: Making the Assessment Connection." *Education Sciences* 4 (4): 247–64. https://doi.org/10.3390/educsci4040247.

Verdake, H, T.D. Mulhern, J Lodge, K Elliott, S Cropper, B Rubinstein, A Horton, et al. 2017. *Misconceptions as a Trigger for Enhancing Student Learning in Higher Education*. The University of Melbourne.

Villarroel, Verónica, David Boud, Susan Bloxham, Daniela Bruna, and Carola Bruna. 2020. "Using Principles of Authentic Assessment to Redesign Written Examinations and Tests." *Innovations in Education and Teaching International* 57 (1): 38–49. https://doi.org/10.1080/14703297.2018.1564882.

Vogel, Susanne, and Lars Schwabe. 2016. "Learning and Memory under Stress: Implications for the Classroom." *Npj Science of Learning* 1 (1): 16011. https://doi.org/10.1038/npjscilearn.2016.11.

Weekes, Nicole, Richard Lewis, Falgooni Patel, Jared Garrison-Jakel, Dale E. Berger, and Sonia J. Lupien. 2006. "Examination Stress as an Ecological Inducer of Cortisol and Psychological Responses to Stress in Undergraduate Students." *Stress* 9 (4): 199–206. https://doi.org/10.1080/10253890601029751.

Winstone, Naomi E., and David Carless. 2020. *Designing Effective Feedback Processes in Higher Education: A Learning-Focused Approach*. London ; New York: Routledge, Taylor & Francis Group.

Wise, Steven L. 2009. "Strategies for Managing the Problem of Unmotivated Examinees in Low-Stakes Testing Programs." *The Journal of General Education* 58 (3): 152–66.

Wise, Steven L, and Christine E Demars. 2005. "Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions." *Educational Assessment*.

Wolf, Lisa F., and Jeffrey K. Smith. 1995. "The Consequence of Consequence: Motivation, Anxiety, and Test Performance." *Applied Measurement in Education* 8 (3): 227–42. https://doi.org/10.1207/s15324818ame0803_3.

Wong, Hwei Ming, Dennis Kwek, and Kelvin Tan. 2020. "Changing Assessments and the Examination Culture in Singapore: A Review and Analysis of Singapore's Assessment Policies." *Asia Pacific Journal of Education* 40 (4): 433–57. https://doi.org/10.1080/02188791.2020.1838886.

Yorke, Mantz, Paul Bridges, and Harvey Woolf. 2000. "Mark Distributions and Marking Practices in UK Higher Education: Some Challenging Issues." *Active Learning in Higher Education* 1 (1): 7–27. https://doi.org/10.1177/1469787400001001002.

Zhang, Zhihong, Hai Su, Qiang Peng, Qing Yang, and Xiaoshu Cheng. 2011. "Exam Anxiety Induces Significant Blood Pressure and Heart Rate Increase in College Students." *Clinical and Experimental Hypertension* 33 (5): 281–86. https://doi.org/10.3109/10641963.2010.531850.

Melbourne Centre for the Study of Higher Education
Level 1, Elisabeth Murdoch Building, Spencer Road
The University of Melbourne
Victoria 3010 Australia
melbourne-cshe@unimelb.edu.au

melbourne-cshe.unimelb.edu.au