FASEB BioAdvances | WILEY

PERSPECTIVE

Considerations and strategies for effective online assessment with a focus on the biomedical sciences

Karen Mate | Judith Weidenhofer

School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, NSW, Australia

Correspondence
Karen Mate, School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan, NSW, Australia.
Email: karen.mate@newcastle.edu.au

This article is part of the Biomedical Education Special Collection.

Abstract

The COVID-19 pandemic in 2020 caused many universities to rapidly transition into online learning and assessment. For many this created a marked shift in design of assessments in an attempt to counteract the lack of invigilation of examinations conducted online. While disruptive for both staff and students, this sudden change provided a much needed reconsideration of the purpose of assessment. This review considers the implications of transitioning to online assessment providing practical strategies for achieving authentic assessment of students online, while ensuring standards and accountability against professional accrediting body requirements. The case study presented demonstrates that an online multiple choice assessment provides similar rigor in assessment to invigilated examination of the same concepts in human physiology. Online assessment has the added benefit of enabling rapid and specific feedback to large cohorts of students on their personal performance, allowing students to target their weaker areas for remediation. This has implications for improving both pedagogy and efficiency in assessment of large cohorts where the default is often to assess basic recall knowledge in a multiple choice assessment. This review examines the key elements for implementation of online assessments including consideration of the role of assessment in teaching and learning, the rationale for online delivery, accessibility of the assessment from both a technical and equity perspective, academic integrity as well as the authenticity and structure of the assessment.

KEYWORDS
academic integrity, authentic assessment, equity, physiology, undergraduate

1 | INTRODUCTION

Teaching and learning in higher education has become increasingly engaged with the online environment over recent years, however, the sudden requirement for online learning and assessment brought on by the COVID-19 pandemic led to the rapid adaption of assessment practices to online delivery. Unlike higher education institutes that have worked within an online distance education model for many years, most institutes were inadequately resourced for the online assessment environment. It has been a major challenge moving from the previously

described "patchy" adoption of technology assisted assessments[1] to rapid wide-scale implementation, while maintaining the confidence of students, staff, employers, and accreditation bodies.

Assessment was defined by Huba and Freed[2] as "the process of gathering and discussing information from multiple and diverse sources in order to develop a deep understanding of what students know, understand, and can do with their knowledge as a result of their educational experiences; the process culminates when assessment results are used to improve subsequent learning." It is an integral element of the teaching and learning process regardless of the discipline, content, and mode of delivery. Indeed it has been said, "Students can, with difficulty, escape the effects of poor teaching, they cannot ....escape the effects of poor assessment".[3] While online assessment has the potential to enhance the teaching and learning process, both practically (to manage distance education, increasing class sizes, and staff workload) and pedagogically (to provide continuous feedback to both students and staff on progress toward learning goals), it presents challenges for academic integrity and student equity. Effective implementation of online assessments therefore requires careful consideration of the role of assessment in teaching and learning, the rationale for online delivery, accessibility of the assessment from both a technical and equity perspective, academic integrity as well as the authenticity and structure of the assessment. This paper provides a review of these areas, summarizes strategies for effective summative assessments online and presents a case study of an online summative assessment in a large first year human physiology course.

2 | ROLE OF ASSESSMENT

Assessment is an integral part of the educational process; it promotes learning and confirms that students have achieved the learning outcomes of the course. In order for both of these requirements to be satisfied, assessment practices need to be aligned with the curriculum and teaching methods, and make use of both formative and summative tasks.[4] Assessment also serves as a motivator of student learning, a number of lower stakes assessments can be used to provide short-term goals for students. However, inappropriate assessments can also lead to a demotivating effect and an excessive quantity of assessment can be overwhelming.

Formative online or digital assessments have been used extensively in many disciplines, including physiology, anatomy, biochemistry, and others, since the introduction of learning management systems (LMSs) to tertiary education.[5] The key feature of formative assessment is that information is released or fed back to the learner to help identify areas of strength and weakness and motivate them to improve their learning and future performance. For the instructor it allows the identification of misconceptions or gaps in learning across the cohort, and to reflect on their own practice. Although a crucial element of the leaning process, feedback practice in higher education is an area of dissatisfaction for both students and staff.[6] The online environment is well suited to a variety of asynchronous formative assessments, that can be used as desired by students to gain feedback on their learning and by staff to monitor student engagement and progress. It has been used in this way in online-based higher education facilities and also within traditional on campus modes of study.

Summative assessments, including examinations, play an important role in ensuring students have factual knowledge, technical proficiencies, communication, and higher order cognitive skills. In the context of biomedical science; students studying medicine, nursing, pharmacy, physiotherapy, podiatry, medical radiation science, speech pathology, occupational therapy, and other allied health professions have accrediting bodies (like many vocational programs) that require a satisfactory means of demonstrating students have met key learning outcomes and standards, with summative assessments performing this role. One of the most fundamental principles of summative assessment in this context is that it closely aligns with learning objectives. Bloom's taxonomy[7] provides a hierarchical framework for development and classification of learning objectives and assessments based on six levels of intellectual activity: knowledge, comprehension, application, analysis, synthesis, and evaluation. With careful attention to this framework, an appropriate balance of questions and activities that address lower, intermediate, and higher order cognitive levels can be used to assess and demonstrate key knowledge, proficiencies, and skills.

3 | RATIONALE FOR ONLINE ASSESSMENT

A recent report on the future of assessment in universities argued that technology could be used to make assessments more authentic, accessible, secure, efficient, and effective.[8] There are both efficiency and pedagogical reasons for the introduction and increasing role of online assessment in higher education,[9,10] however, these are balanced by practical challenges and risks, especially for summative assessments.

3.1 | Pedagogy

One of the most compelling pedagogical reasons for online assessment is the opportunity to provide immediate

and meaningful feedback that aligns with the principles of good feedback practice. Good feedback clarifies the standards and criteria of good performance; assists development of reflective learning; provides quality information to both students and teachers to facilitate learning and teaching; encourages positive motivation and dialogue around learning; and gives opportunities to progress from the current to desired level of performance.[11] While marks and grades are an important aspect of feedback, they alone do not enhance learning and can actually hinder learning and decrease motivation.[9] Indeed, the specificity of feedback is a major challenge faced by both students and staff; students desire more detailed and less generic feedback, while staff feel that inadequate workload allocations and lack of scalability limit their ability to provide personalized feedback.[9] Online testing provides an opportunity to provide not only a grade, but also specific feedback about correct responses and the reasoning involved. Formative multiple choice (or similar automatable marking question type) activities and tests, when aligned with the principles of good feedback, can be an efficient and effective online strategy to support student learning and autonomy.[11]

Advances in machine learning have led to the development of adaptive learning platforms (using learning analytics). These platforms create a more personalized assessment system that can monitor and respond to user input.[12] The fast feedback and adaptation to the progression of the student, coupled with direction to electronic resources to remediate areas of misunderstanding, is particularly suited to formative assessment to support student learning. Furthermore, these platforms provide a useful means to monitor and evaluate learner engagement with the resources; recording activity and progress, to facilitate dialogue between the student and instructor.

3.2 | Efficiency/Practicality

In designing any assessment task, be it summative or formative, it is essential that it is practical for both students and staff. In particular, time needed for completion and grading,[13] must be balanced with achieving a task that provides information on achievement of a learning objective. The development of the World Wide Web and internet in the 1990s, and subsequent widespread adoption of digital LMSs in higher education in the early 2000s, provided a previously unseen ease of delivering distance teaching, learning, and assessment, which led to what is now known as online learning. Online assessment, also referred to as e-assessment, technology-assisted assessment or computer-based testing, represents perhaps the most challenging component of online higher education.

Global technological advances also triggered changes in the tertiary education sector, in particular a rapid expansion in degree options and student numbers. Alongside a concomitant decrease in resources, this led to increased student to staff ratios and staff teaching and assessment workloads. To overcome this many staff shifted assessments toward high-throughput marking styles. Certain types of assessments created, edited, and deployed using a course LMS (e.g., Blackboard, Moodle, and Canvas) can be marked automatically, and provide immediate specific feedback to students on their answers including a short explanation of the correct and incorrect answer options. While designing and writing online assessments that develop and/or measure different cognitive levels is challenging and time consuming, automated marking, and feedback of some question types (e.g., multiple choice, true/false, matching, and short response) increases efficiency for both student and instructor. Consequently, these types of assessments are regularly used in large first year undergraduate courses across the globe for both formative and summative needs. This has also been recognized by textbook publishers such as Pearson Education, Wiley, and McGraw-Hill Education that have developed comprehensive online packages of learning and assessment materials that integrate within the LMS. These packages include extensive test banks of different question types that be used as the basis for creation of both formative and summative assessments.

As teaching and learning has increasingly (and then suddenly during the COVID-19 pandemic) moved to online delivery, the inclusion of online assessment has alleviated the validity associated with the previous disjunction between teaching and assessment modes with e-based learning.[9] Despite this, there is still much debate, both globally and locally, on whether online assessments, particularly examinations, offer the same academic integrity as a traditional on campus paper assessment.

4 | ACCESSIBILITY OF ASSESSMENT

4.1 | Technical

A 2016 study of first year tertiary student attitudes found that information and communications technology (ICT) infrastructure and reliable connectivity were significant barriers to successful completion of online examinations at that time.[14] From the institutional perspective, access to reliable systems that can manage synchronous delivery to large numbers of students and technical support for staff and students is essential. Contingency plans for dealing with technical failures are vital when planning

and delivering online assessments, as demonstrated by an account of a central server failure during a large undergraduate summative test.[15]

Higher education providers have a responsibility to ensure that their virtual learning and assessment environments are usable and accessible. Commonly used LMSs are designed in accordance with the internationally recognized Web Content Accessibility Guidelines, so they can be utilized by all students regardless of their needs and preferences. Delivery of online assessments within a compliant LMS and in keeping with the general accessibility principles applicable to online learning[16,17] will benefit all learners, however, instructors should also design their course site with usability in mind. The mark that a student achieves in an online assessment should be a reflection of their achievement against course learning goals, not their IT skills. Formative and summative assessments, grades, and feedback should all be easy to find on the course site, which may be best achieved through a course site designed with both functional and chronological elements.[18]

4.2 | Equity

A strong motivation for students' own learning is the enjoyment they receive from having opportunities to demonstrate their learning success.[13] All students, regardless of their individual background and circumstances, should be provided with the opportunity to demonstrate their learning. Online assessment can potentially exacerbate educational and social disadvantage due to differences in access to, and literacy in digital technologies associated with age, gender, and socioeconomic factors.[1] In addition, there are other less tangible disadvantages to online assessment, which may be more relevant to particular students. Small test-mode effects have been observed, with high performing students benefitting most from a shift away from paper-based to online assessments,[19] possibly due to differences in cognitive workload.[20] Others report no increase in cognitive load associated with online testing in a tertiary setting.[21] A study designed to separate online testing environment effects from cheating effects found that exam performance was negatively affected by the online testing environment, which was offset by an increased propensity to cheat.[22] Several negative effects of the online testing environment were identified including disadvantages due to greater distraction, technical difficulties, and the inability to seek clarification for any ambiguous questions.

While studying and completing assessments "online" presented a degree of flexibility in learning that may be conceived as an advantage to those students with work and family caring responsibilities, the sudden closure of schools and childcare centers during the COVID-19 pandemic created increased distraction.[23] In some cases, older siblings were also affected by lost "quiet" study time with younger siblings not attending school; there was increased competition in the home for use of technology and a very high demand on internet bandwidth. While all these inequities exist and likely many more, such as an exacerbation of gender equality in caring responsibilities during the COVID-19 pandemic, it is important to be aware that students experience different online testing environments depending on their personal circumstances.

There is also a diversity of technological preferences and proficiencies among all students, including those younger students, who are commonly referred to as "digital natives".[24] Interestingly these "digital natives" often lack capacity to trouble shoot or cope when technology fails to work as anticipated or something varies slightly from expected. It is these students who appear to fare worse in these circumstances, possibly due to a lack of having to work out technology on their own and becoming over reliant on what they are "taught" at school. Equity in access to suitable technology and support for completing varied online assessments became especially relevant in the rapid shift to online assessment due to COVID-19. Limitations reported by students included lack of home Wi-Fi, failure of audio and/or visual connections, and sharing computers with other family members.[23] Many students rely on university or library computers, which became more intensively used and/or were in restricted access areas due to social distancing requirements. This situation was further exacerbated by general shortages of electronic and computing equipment, limited access to shops, and delayed delivery times.

Reasonable adjustments to online assessments may be required to address inequities due to physical or mental limitations. Many higher education students, for example, have some form of clinically diagnosed anxiety. These students are often allowed additional time and or other adaptations to the usual paper-based exam to allow them to focus, lessen the chance of an acute episode of anxiety, and achieve deeper learning.[13] A recent survey of students from a range of health disciplines revealed that 38% of female and 25% of male respondents experienced more stress with remote online exams than face-to-face exams.[25] The main stressors reported were exam duration, navigation mode (backtracking permitted or not; questions presented one at a time or all at once), and technical problems.[25] It is likely therefore, that online assessments increase anxiety further in those with underlying anxiety disorders and as such, the existing adjustments for testing for these students may be inadequate and create an equity concern. Further research in this area is clearly required

to determine if standard adjustments for a variety of medical conditions that students have, are as equally appropriate for online assessment as they are for paper and oral assessments. This includes consideration by the assessor of the layout and adjustability of text and image size and color and how this may be affected by technological limitations.

5 | ACADEMIC INTEGRITY

The prevalence of cheating, or willingness to cheat, on graded assessments among tertiary students has increased steadily over a number of years, with estimates ranging from 9% to as high as 90%.[26] There is a strong perception, and some evidence that the growth in use of online assessments presents a threat to academic integrity as they may provide increased opportunity for student cheating compared to traditional invigilated face-to-face exams. Since 2011, there has been a steep rise in the number of students undertaking higher education online; in 2016 more than a third of Australian university students enrolled in at least one online course.[27] The issue of academic integrity of assessment was becoming a priority area with the number of students completing a substantial component of their studies off campus predicted to increase rapidly, even before the COVID-19 pandemic forced a large proportion of courses online in 2020.[28] While academic dishonesty is a serious concern for the capacity of any graduate to undertake future work, within the health disciplines the potential consequences for patient care could result in serious injury or death and as such this is a priority concern for academics responsible for educating these students.[29] Indeed there has been a recent call for institutions to annotate summative test results of medical students as "in-person proctored", "online proctored," or "unproctored" in order to maintain transparency and accountability.[10,30]

Academic dishonesty in online assessments without direct supervision can take several forms[31]:

- *Misrepresentation of identity*, in which a third party completes individual assessments or even the entire course for the student. The third party may be known to the student (e.g. a family member or friend) or an unknown "work for hire" arrangement.
- *Plagiarism*, defined by our institution (The University of Newcastle) as "presenting the words or ideas of someone else as your own without giving credit to the original author."
- *Manipulation of technology* to gain advantage can take several forms; students may deliberately crash or break their internet connection to gain extra time or the opportunity to re-take an assessment, or find technical

loopholes to access pre-set answers or other students submitted work.

- *Collaboration or collusion* with other students taking the course to share information and answers while completing the assessment.
- *Deception* or breaking the agreed conditions for completion of an assessment, such as referring to notes or other sources of information during an online test. Although this particular type of behavior is normatively described as cheating, 46% of face to face and 71% of online students believe that referring to existing notes is not cheating at all or only trivial cheating.[26]

To ensure integrity of assessment, tasks should generate clear evidence that the work, whatever its nature, has been produced by the candidate.[13] Different approaches are required to address the different types of cheating and assessments in the online environment. Measures such as including verification of the test taker, plagiarism detection software and supervision of monitoring of test conditions can directly reduce cheating, whereas other approaches such as use of authentic assessments can reduce both the opportunity and motivation to cheat.

5.1 | Authentication and Security

One aspect of reducing cheating is authenticating the user and monitoring the activity of the user during the assessment. A number of proctoring solutions, including LMS add-ons (e.g., Respondus within Blackboard, Waevaer within Moodle), and other services (e.g., ProctorU), use video, keystrokes, fingerprints, and the like to identify the user during the test. Some of these software solutions provide algorithms that monitor eye movement and other motion to determine if the user is potentially utilizing "off screen" notes, while others track activity arising from the device such as use of internet browsers. In this way, these services allow academics to overcome many of the issues raised above regarding cheating. However, they raise a secondary issue around security and protection of user privacy. In particular, use of video and monitoring of device activity allow for unintended information to be shared and/or unethically obtained. Furthermore, large amounts of data are collected by these softwares in order to undertake their algorithms and this can also entice unethical use of student specific data. An alternative approach is the use of post-hoc assessments, such as an oral examination or *viva voce* to validate results. However, this requires either extensive time commitment to interview all students or creates equity issues around which students are interviewed and what to subsequently do with information on a subset of the cohort. Alternatively, rather than query the

student on content knowledge using post-hoc oral examination, students could be queried on their approach to the assessment. This may be quite stressful for some students if they feel they are being "accused" of cheating and could affect their ability to recall their approach and explain their thinking.

5.2 | Proctored vs Unproctored

There have been reports that students perform better in non-proctored exams than proctored exams. A recent comparison of fourth year medical students results from an open-book online examination, necessitated by the COVID-10 pandemic, with results from traditional examinations found significant differences.[32] The mean scores for the multiple choice and essay questions were higher than for closed-book examinations. However, the scores for short answer questions were lower. Similarly, others report increased completion times and lower scores when online tests were proctored.[33,34] The authors indicated that the online and closed-book examination questions were of the same difficulty, but did not provide any information on the cognitive level of questions,[33,34] while one specifically mentioned that the exam required a high degree of memorization.[32]

Others have reported no significant difference between proctored or invigilated tests and non-proctored tests.[35-37] Key features of the non-proctored tests in these studies include the use of a lockdown browser and the inability to backtrack,[35] low stakes assessment,[36] and a commitment to authentic assessment.[37] As such, the determining factors of using proctored assessment or not appears to be similar to those for determining the use of open-book and closed-book examination in the traditional paper-based invigilated setting. Regardless, the most appropriate choice should depend on the learning outcomes and skill set being assessed.[38]

5.3 | Authentic Assessment

As with all assessments, online assessment must meet a measure of effectiveness: that is the assessment tasks should be designed to encourage good quality, "deep" approaches to learning in the students. This is true regardless of whether the learning outcome is for the acquisition or for the application of knowledge. Assessment can be described within three paradigms[39]: (i) assessment as measurement, (ii) assessment as procedure, and (iii) assessment as inquiry. Another, perhaps simpler way of presenting these three paradigms is as follows: (i) assessment *of* learning, (ii) assessment *for* learning, and (iii)

assessment *as* learning.[13] Authentic "inquiry" assessments contextualize learning and include consideration of the complexities and ambiguities of operating in the real-world situation. Assessments constructed in this way can provide recognition and evidence of student performance in tasks requiring higher order cognitive skills.

5.4 | Validity and Reliability

A number of principles relevant to validity and reliability of online assessment have been elucidated by various authors and regulatory bodies.[13,40] The underlying principle to ensure an assessment is authentic is that it be designed such that it accurately reflects the assessed learning outcomes.[41] As such an assessment is considered to have validity if it can meet expectations of both the assessors and the students. Validity can be considered in regards to content: ensuring the assessment tasks are assessing the stated learning outcomes, and fairness; is satisfactory performance achievable in the modality used and the time allowed, which is typically only able to be measured after completion.[42] The issue of fairness is also confounded by conflicting expectation of academics and students, particularly when introducing online assessment into a course where it was not previously used. Information, guidance, rules, and regulations on assessment should be clear, accurate, consistent, and accessible to all staff, students, practice teachers, and external examiners.[13,40]

The online environment without proctoring makes it difficult to generate an authentic assessment for the acquisition of knowledge that is also fair. The application of knowledge, however, is far easier to assess in an online assessment while maintaining fairness. It is also appealing for academics to assume that general underperformance in an online assessment is the result of student expectation that they will be able to use a multitude of resources to assist them in answering the questions and that they will have ample time to do this, rather than that the assessment was indeed not fair in terms of the restrictions placed upon it in regards settings such as time and backtracking.

An intrinsic requirement of assessment design is that the task should generate comparable marks across time, across markers, and across methods. Objective marking criteria and electronic marking can improve reliability, making online assessment an attractive option. A common approach to reducing collusion in online examinations is the use of question pools; in this instance it is imperative that the selection of questions is structured to ensure fairness during the random distribution of questions. This requirement necessitates the development of a high number of small pools of questions to ensure that an

equitable selection of questions (of different topics, difficulty, and cognitive level) is presented to each student.[43] Alternatively question variants with different numerical values could be supplied to individuals at random, but again these must represent equal difficulty and are only achievable for certain topics.

5.5 | Multiple choice question assessments

The myth that multiple choice questions (MCQs) only assess lower cognitive levels and promote surface learning has now been well and truly debunked; well-constructed MCQs can objectively measure learning achievement across higher cognitive levels such as application, analysis, and problem solving.[44-46] While constructing a well-designed MCQ is challenging, there are numerous generic guides available,[47,48] as well as more specialized frameworks for construction of MCQs targeting different Bloom's taxonomy levels in specific disciplines.[44,45,49] While pictures are easily incorporated into traditional paper-based assessments, additional features such as animations, audios, and videos can increase the authenticity of questions that can be constructed compared to traditional paper-based assessments.[42] When done well, MCQ-based assessments can fulfil the parameters of authentic, valid, and reliable assessment, provide timely feedback to students and instructors and improve efficiencies in large undergraduate classes.

6 | STRATEGIES FOR ONLINE DELIVERY OF AUTHENTIC, VALID, AND RELIABLE MULTIPLE CHOICE QUESTION ASSESSMENTS

A number of authors[50,51] have suggested "control procedures" to help minimize cheating in online exams, many of which are specifically relevant to online multiple choice assessments in the absence of proctoring. These include:

- *Academic integrity or code of conduct statement*: Enhanced awareness and knowledge of expectations of academic integrity for students including the problems, implications, and potential consequences of academic dishonesty can help change attitudes and reduce engagement in cheating behaviors.[29] Inclusion of an academic integrity declaration that must be acknowledged or digitally signed in order to access each online assessment can inform and remind students of general academic integrity standards as well as the specific conditions for the assessment.

- *Timing*: Scheduling a single set time to complete the assessment will prevent students from collaborating and completing it sequentially. The effectiveness of this approach can be increased by making the assessment link available within a narrow time window to force all students to commence and therefore complete the assessment simultaneously. Restricting the time allowed for completion of the exam so it is sufficient for students to provide thoughtful answers but not to research or find answers is another timing-related strategy that is particularly useful when assessing knowledge, recall, and interpretation. In this case, the assessment should be set to auto-submit when the allotted time expires. Individual adjustments may be required for students with a disability.

- *Content*: We have already defined assessment to include a measure of what students "know, understand, and can do with their knowledge, as a result of their educational experiences".[2] A modified Bloom's taxonomy[7] based approach to construction of MCQ assessments with a greater focus on higher order cognitive skills (application, analysis, synthesis, and evaluation) than lower order skills (knowledge and comprehension) can also address academic integrity. Answers to questions requiring higher cognitive levels cannot be looked up in notes, textbooks, or found by online searches. There is a tool available to guide writing and classifying questions on biology-related topics using Bloom's taxonomy.[49] Images and graphs are an effective way to create higher order questions as long as they are not identical to those used previously in teaching activities.[52] Any assessment delivered online should subsequently be considered to be freely available in the public domain and therefore questions should be changed regularly or modified between cohorts. Algorithmic test banks that present variations of questions by changing question parameters at each implementation to create personalized exams have been used for particular topic areas such as mathematics,[53] but are achievable in other areas, including the health sciences.

- *Other parameters*: The presentation of questions one at a time, and randomized for each test taker will minimize the opportunity for synchronous collusion and completion of the assessment. Randomizing the answers to each question, and preventing students from backtracking to questions they have already answered are additional measures that limit opportunities to cheat. The incorporation of images into the question stem can also make it more difficult for students to copy and paste key words into search engines. Any image files incorporated into an online assessment should be given an uninformative name, as it may be accessible to students. It is also important critical to

consider the implications of this approach on the cognitive level of the question.

7 | ALTERNATIVE TYPES OF SUMMATIVE ONLINE ASSESSMENT

There are a variety of other assessment types that are well suited to the online environment, and have the advantages of automated marking and immediate feedback to students including extended matching exercises, fill in the gap (cloze passage) or single word/number answers, labeling or identification of image hotspots, drag and drop labeling, and online simulations. Essentially any type of assessment can be adapted to an online setting including those that address development of skills such as communication or teamwork. In this context, assessments may include short or long answer questions, submission of essays or case studies, participation in discussion forums, wikis, reflections, e-portfolios, as well as individual or group presentations (either live or recorded).

While these types of assessments have potential for individualized instruction and targeting of specific learning needs, they present their own challenges. Some may not be applicable to very large undergraduate classes due to the burden of manual marking and the inability to provide immediate or timely feedback tailored to the individual, which provides an opportunity to correct misunderstandings and misconceptions. Although intuitively we may assume that essays, case studies etc. provide a more thorough and rigorous assessment strategy this is not necessarily the case. A comparative analysis of MCQ and modified essay questions (MEQs) used in assessments of fourth year medical students found that the majority of both question types focused on recall of knowledge, however, MCQs were better at addressing the highest order cognitive skills compared with MEQs.[45] MEQs tend to test knowledge as well as higher cognitive skills, although they do have the advantage of contributing to the development of written communication skills especially if high-quality feedback in this area is provided to students.

8 | CASE STUDY

As the COVID-19 pandemic caused international community "lockdowns", face-to-face courses at many higher educational institutes were rapidly transitioned to fully online delivery. In our case, this impacted the delivery of classes, as well as the major assessment items (invigilated mid-semester and final exams) for an introductory human physiology course undertaken by first year students from multiple degree programs (biomedical science, nutrition and dietetics, pharmacy, physiotherapy, podiatry, speech pathology). Traditionally, the course is delivered face to face across two campuses, with approximately 500 students enrolled at the main campus and 100 students enrolled at a regional campus about 100 km away. There is a single course Blackboard site that serves students from both campuses. Our largest challenge was to transition invigilated assessments to an online (non-proctored) format, while ensuring and demonstrating the academic integrity of the course to satisfy the accreditation requirements of the multiple degree programs that this course services. This case study describes our approach to balancing student equity and quality assurance in the mid-semester assessment that occurred at the time of intensive restrictions to limit the spread of COVID-19.

8.1 | Equity considerations

Our university has a relatively high proportion of students from low socioeconomic backgrounds, the majority of whom have gained entry to their program via alternative pathways for non-school leavers. A number of students were therefore affected by the closure of childcare centers and schools, and expressed concern about completing the mid-semester assessment at the originally scheduled times (while caring for young children). Although we had already decided to adhere to the original course timetable, including separate exam dates (on consecutive days) for the mid-semester assessment at the two campuses, this did not address the issue of students caring for small children. To ensure that all students had the opportunity to complete the exam under suitable conditions an additional evening time slot was scheduled, which meant that there were three separate exam time slots over 2 days that students could choose from. Three versions of the 30 MCQ mid-semester assessment were created; there were 12 questions in common, 9 variant questions (with different values in each sitting, as described below), and 9 unique questions. There was a single sitting for the end of semester assessment, as many of the previous issues relating to social restrictions had been lifted by that time.

Considerations and adjustments to the time allowed to complete the exam were given to students with a disability. At our university students with a disability are able to register with the health service to develop a reasonable adjustment plan that specifies adjustments that are required to support their learning including font size and color text to voice technology and additional time required to complete assessments.

8.2 │ Academic integrity considerations

Online delivery and the multiple opportunities for completion of the mid-semester assessment were identified as potential threats to academic integrity that required control and monitoring. We therefore implemented several strategies to maximize within-test and between test academic integrity by requiring students to review and agree to an academic integrity statement before starting the assessment; setting a 40-min time limit for completion with auto submission when that time elapsed; presentation of one question at a time with no option to return to questions they had already answered and saved (e.g., no backtracking); presentation of questions and answer options in a random order; and making the link to the assessment available within a limited 30-min time window to ensure a nearly synchronous start for all students. This combination of settings was chosen to limit opportunities for collaboration and cheating. Question variants (Figure 1) were used to assess if students had gained access to the questions or the answers from a previous sitting. As such the questions were varied so that all the answer options remained the same, but the correct answer was different.

An additional measure to ensure integrity and improve the quality of the assessment was to adjust the questions toward assessing higher order cognition. The answers to higher order and discipline integrated questions cannot be easily found in notes, textbooks, or by searching online and are therefore a more robust indicator of what "students know, understand, and can do with their knowledge".[2] Questions were classified into three categories using a modified Bloom's Taxonomy scale: Level 1 (L1) questions were simple recall questions, level 2 (L2) questions required some interpretation, and level 3 (L3) questions involved application and or analysis of key concepts. For the 30-question assessment, the number of L1 questions was reduced from 13 (2019; invigilated) to 7 (2020; online) and the number of L3 questions was increased from 6 (2019; invigilated) to 12 (2020; online). The number of L2 questions remained relatively constant and comprised approximately 40% of the exam (Figure 2A). Table 1 shows an example of a question that was changed from Bloom's L1 in the 2019 invigilated exam to L3 in the 2020 online exam. The L1 question requires only recall of the key features of the different fluid compartments within the body, whereas the L3 question requires an understanding of the relevance of these differences and what it means for cell function. These questions were written and classified according to Blooms level by subject matter experts, however, recent research has found that students may approach questions differently than intended.[54]
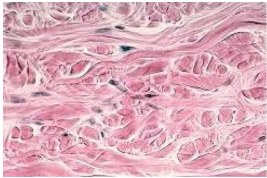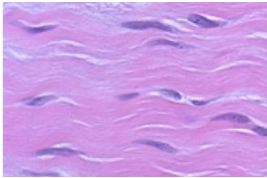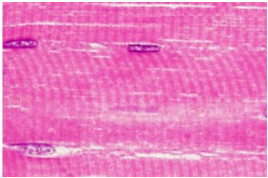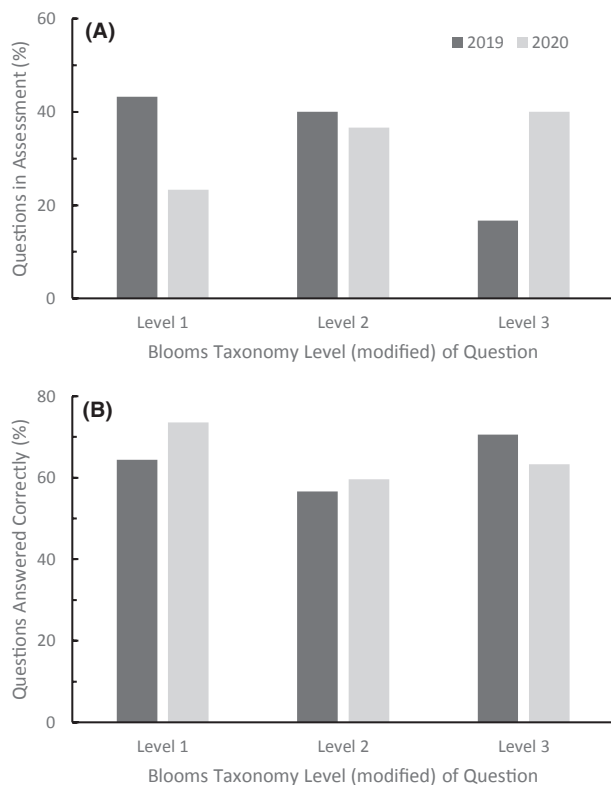
Stem Variation 1	Stem Variation 2	Stem Variation 3	Answer options
An element has 2 neutrons and 12 electrons in a cloud orbiting the nucleus. What is the mass of one mole of this element? Correct response B	An element has 12 neutrons and 28 electrons in a cloud orbiting the nucleus. What is the mass of one mole of this element? Correct Response D	An element has 2 neutrons and 10 electrons in a cloud orbiting the nucleus. What is the mass of one mole of this element? Correct response A	A. 12 grams B. 14 grams C. 26 grams D. 40 grams
During digestion disaccharides are converted to monosaccharides by the _____. Correct response A	An anabolic reaction involving disaccharides and monosaccharides would involve the _____. Correct response C	The conversion of dipeptides to amino acids involves the _____. Correct response A	A. addition of a water molecule between each two units B. addition of a carbon atom between each two units C. removal of a water molecule between each two units D. removal of a carbon atom between each two units
Ehlers-Danlos Type VI is an autosomal recessive condition characterized by high risk of retinal detachment. When one parent has detached retinas, and the other is heterozygous for the trait, what is the possibility that the second child will have Ehlers-Danlos Type VI? Correct response C	Polydactyly is an autosomal dominant condition characterized by the presence of extra finger(s). When both parents are heterozygous for the trait, what is the possibility that the first child will have 5 fingers on each hand? Correct response B	Ehlers-Danlos Type VI is an autosomal recessive condition characterized by high risk of retinal detachment. When one parent has detached retinas, and the other is homozygous for the trait, what is the possibility that the second child will have Ehlers-Danlos Type VI? Correct response D	A. 0% B. 25% C. 50% D. 100%
The tissue in the image is best described as: Correct response D	The tissue in the image is best described as: Correct response C	The tissue in the image is best described as: Correct response A	A. Skeletal muscle B. Loose connective tissue C. A tendon or ligament D. Dense regular connective tissue

FIGURE 1 Design of "calibrator" multiple choice questions for multiple offerings of the same online exam

8.3 | Findings

Analysis of the results from the three online offerings of the assessment showed that a higher mark was obtained by students in the first sitting (mean, 20.8) compared to the second (mean, 18.7; $p < 0.01$, ANOVA with post-hoc Tukey HSD) and third (mean 19.7; $p < 0.05$, ANOVA with post-hoc Tukey HSD) sittings. Students in the first sitting may have been better prepared, more confident with the content and presumably had fewer

FIGURE 2 Comparison of examination content and student performance in a paper-based invigilated format in 2019 and a non-invigilated online format in 2020 for a first year human physiology course. (A) The proportion of questions classified as modified Bloom's level 1, 2, or 3. (B) The proportion of correctly answered questions based on their Bloom's level 1, 2, or 3 classification

impediments to study (e.g., young children), however, this was not specifically examined. Overall we were satisfied that the results were generally consistent across the three sittings, with no obvious signs of cheating or collusion. Although not directly comparable, the average mark for the 2020 online mid-semester assessment (19.3 of 30) did not differ significantly from the 2019 invigilated assessment (18.6 of 30; $p = 0.3$, t-test) and is consistent with results over several years. As we expected, the performance in L1 recall questions was higher in the 2020 "open book" online exam, but the performance in the L3 application/analysis questions was lower (Figure 2B). This may be due to the test settings that prevented backtracking; questions had to be answered as they were presented so students did not have an opportunity to go back or spend longer thinking about the more challenging questions. Indeed, many students reported that the inability to backtrack was the most challenging aspect of online assessment, as they could not use their normal examination strategy of returning to difficult questions. The randomization of question order may have exacerbated this situation for some students, as they could potentially be presented with the most difficult questions at the start of the exam affecting their confidence and time management.

The built in countdown timer in the Blackboard-based online assessment was another stressor reported by our students that is not present in a traditional paper-based exam. We have also observed that time management is poor in online exams; 5.4% of students did not reach the final question, and 3.0% did not answer the final three questions (10% of the exam), whereas more than 99% of students completed all questions in previous paper-based versions of the exam. Students also report a high rate of misreading questions off screens, in part due to not being able to underline or highlight keywords within the question. Additionally, students when being examined online are less inclined to use pen and paper to assist them with their problem solving. While these issues are primarily the concern of the individual

TABLE 1 Questions on fluid compartments within the body that written to test learning at a modified Bloom's taxonomy level 1 and level 3

Bloom's level	Question stem	Answer options
1	Which body fluid compartment contains high levels of K+, large anions and proteins?	A. Intracellular fluid (correct answer) B. Interstitial fluid C. Plasma D. Both plasma and interstitial fluid
3	An artificial cell contains 25 potassium ions and 35 protein anions. The surrounding solution contains 30 chloride ions and 30 sodium ions. What is the membrane potential difference?	A. −10 (correct answer) B. 10 C. −35 D. −60

student, the academic still maintains some responsibility to ensure that students are aware they are permitted to do so, particularly where online assessment is being used as a major component of an otherwise face-to-face course. In these circumstances students are often lost in the online component of the task rather than focussing on demonstrating their capacity against learning outcomes.

Overall the results of the changes made compared to the overall performance showed that with careful planning and adjustments an "open book" online MCQ assessment can provide similar rigor and discriminating power as a "closed-book" invigilated assessment. Our key recommendations to ensure this are to include a higher proportion of questions that assess higher order learning in addition to adjusting test settings and introducing "new" or variable questions delivered in a random fashion to limit opportunities for online collaboration and sharing of answers. With these modifications we were able to offer the test at several different scheduled times so that students were not disadvantaged due to their additional childcare responsibilities or other commitments during the COVID-19 lockdown period. While this lockdown style of learning will lift, considerations of this nature remain for institutes that offer flexible modes of study in order to support students with competing commitments. Further analysis and delivery of tests with alterations to settings is required to determine if all the restrictions imposed are essential to ensure integrity. Of particular note is prevention of backtracking; our data suggested this may have negatively impacted performance in higher order thinking questions in online delivery. It is therefore important to consider the requirement for this feature if the bulk of a test is questions of this nature.

An unexpected benefit of the online test delivery was that we could access additional performance analytics and provide all students with personalized feedback on their performance in the specific topic areas that were tested without releasing the actual questions. The individual student results were available to the instructors immediately in a format that allowed for rapid dissemination. As a result, once quality control was completed, students were emailed (using Microsoft mail-out capabilities) their individual performance on the specific topics assessed by the MCQs. This allowed all students a timely opportunity to remediate their understanding of these topics prior to moving too far ahead in the remaining course material. Previously, it was not feasible to offer such feedback, with students only receiving their individual score and then much later a discussion of topics that were answered poorly across the cohort. In this way, it was possible to provide a desirable feedback on performance for a summative assessment and also maintain the integrity of the questions for future use.

9 | CONCLUSION

The COVID-19 pandemic caused a major disruption to higher education in 2020–2021, increasing the need for flexibility and adaptability, while accelerating the trend toward online teaching and learning. With this transition comes the need to provide valid and reliable measures of student learning, that are accessible and secure, and ensure academic integrity, but are also equitable for all students. Future research will be required to ensure that rapid adoption of online assessment does not result in unintended poor outcomes for students. As it is likely that many universities will continue a level of online assessment in a face-to-face environment, it is vital that the practices that have been rushed into place are carefully considered for ongoing use. In particular, the potential issues raised here around equity in regard socioeconomic factors as well as disability need intentional investigation and appropriate solutions.

Likewise, there is yet to be a comprehensive analysis of how policies introduced to "maximize academic integrity" during online testing, affect capacity of students to demonstrate their achievement against course learning outcomes. Our case study suggests that not all strategies are necessarily best employed together. Where appropriate, shifting questions toward higher Bloom's taxonomy is likely to counteract the need to significantly reduce the time allowed for completion of an assessment in order to restrict looking up or searching for answers. However, this is not possible in all situations, depending on the course learning objectives. Furthermore, while restriction of backtracking is an attractive means to prevent collusion and sharing of answers even for higher Bloom's taxonomy questions, it also leads to poor time management, additional stress, and therefore potential under performance. Further research in this area is warranted. Finally, the increased attention given to development of authentic, reliable, and equitable assessments in the online environment can only benefit our students and several general aspects of assessment that we have discussed here can also be applied to improvement of on-campus assessment practices when we return to face-to-face learning.

AUTHOR CONTRIBUTIONS

The review was written jointly by KM and JW. Both authors contributed equally to the development of the

assessment used within the case study. KM conducted the analysis, with assistance from JW.

ORCID

Karen Mate ![ORCID] https://orcid.org/0000-0002-4255-7937

REFERENCES

1. Timmis S, Broadfoot P, Sutherland R, Oldfield A. Rethinking assessment in a digital age: opportunities, challenges and risks. *Br Edu Res J*. 2016;42:454-476.
2. Huba ME, Freed JE. *Learner-Centred Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Allyn & Bacon; 2000.
3. Boud D. Assessment and learning: contradictory or complementary? In *Assessment for Learning in Higher Education*. In: Knight P, ed. Kogan; 1995:35-48.
4. Biggs J. Assessment and classroom learning: a role for summative assessment? *Assess Educ*. 1998;5:103-110.
5. Marden NY, Ulman LG, Wilson FS, Velan GM. Online feedback assessments in physiology: effects on students' learning experiences and outcomes. *Adv Physiol Educ*. 2013;37:192-200.
6. Henderson M, Ryan T, Phillips M. The challenges of feedback in higher education. *Assess Eval Higher Educ*. 2019;44:1237-1252.
7. Krathwohl DR. A revision of Bloom's taxonomy: an overview. *Theory Into Pract*. 2002;41:212-218.
8. Pauli M, Ferrell G. *The future of assessment: five principles, five targets for 2025*. JISC; 2020:26 p.
9. Gipps CV. What is the role for ICT-based assessment in universities? *Stud High Educ*. 2005;30:171-180.
10. Boitshwarelo B, Reedy AK, Billany T. Envisioning the use of online tests in assessing twenty-first century learning: a literature review. *Res Pract Technol Enhanc Learn*. 2017;12:16.
11. Nicol D. E-assessment by design: using multiple-choice tests to good effect. *J Further High Educ*. 2007;31:53-64.
12. Forsyth B, Kimble C, Birch J, Deel G, Brauer T. Maximizing the adaptive learning technology experience. *J High Educ Theory Pract*. 2019;16:80–88.
13. Bloxham S, Boyd P. *Developing Effective Assessment in Higher Education: A Practical Guide*. McGraw-Hill; 2007.
14. James R. Tertiary student attitudes to invigilated, online summative examinations. *Int J Educ Technol High Educ*. 2016;13:19.
15. Harwood I. When summative computer-aided assessments go wrong: disaster recovery after a major failure. *Br J Edu Technol*. 2005;36:587-597.
16. Phipps L, Kelly B. Holistic approaches to e-learning accessibility. *Res Learn Technol*. 2011;14(1):69-78.
17. Badge JL, Dawson E, Cann AJ, Scott J. Assessing the accessibility of online learning. *Innov Educ Teach Int*. 2008;45:103-113.
18. Conley Q, Earnshaw Y, McWatters G. examining course layouts in blackboard: using eye-tracking to evaluate usability in a learning management system. *Int J Human-Comput Interact*. 2020;36:373-385.
19. Clariana R, Wallace P. Paper–based versus computer–based assessment: key factors associated with the test mode effect. *Br J Edu Technol*. 2002;33:593-602.
20. Noyes J, Garland K, Robbins L. Paper-based versus computer-based assessment: is workload another test mode effect? *Br J Edu Technol*. 2004;35:111-113.
21. Prisacari AA, Danielson J. Computer-based versus paper-based testing: investigating testing mode with cognitive load and scratch paper use. *Comput Hum Behav*. 2017;77:1-10.
22. Fask A, Englander F, Wang Z. Do online exams facilitate cheating? An experiment designed to separate possible cheating from the effect of the online test taking environment. *J Acad Ethics*. 2014;12:101-112.
23. Kalman R, Macias Esparza M, Weston C. Student views of the online learning process during the Covid-19 pandemic: a comparison of upper-level and entry-level undergraduate perspectives. *J Chem Educ*. 2020;97:3353-3357.
24. Bennett S, Maton K. Beyond the 'digital natives' debate: towards a more nuanced understanding of students' technology experiences. *J Comput Assist Learn*. 2010;26:321-331.
25. Elsalem L, Al-Azzam N, Jum'ah AA, Obeidat N, Sindiani AM, Kheirallah KA. Stress and behavioral changes with remote E-exams during the Covid-19 pandemic: a cross-sectional study among undergraduates of medical sciences. *Ann Med Surg (Lond)*. 2020;60:271-279.
26. Burgason KA, Sefiha O, Briggs L. Cheating is in the eye of the beholder: an evolving understanding of academic misconduct. *Innov High Educ*. 2019;44:203-218.
27. Norton A, Cherastidtham I, Mackey W. *Mapping Australian higher education 2018*. Grattan Institute; 2018.
28. Gamage KA, Silva EK, Gunawardhana N. Online delivery and assessment during covid-19: safeguarding academic integrity. *Educ Sci*. 2020;10:301.
29. Azulay Chertok IR, Barnes ER, Gilleland D. Academic integrity in the online learning environment for health sciences students. *Nurse Educ Today*. 2014;34:1324-1329.
30. Prakash ES, Kacsoh B. Unproctored online summative assessments during the covid-19 pandemic: a plea for transparency. *Acad Med*. 2021;96(2):e8.
31. McGee P. Supporting academic honesty in online courses. *J Educ Online*. 2013;10(1):1-31.
32. Eurboonyanun C, Wittayapairoch J, Aphinives P, Petrusa E, Gee DW, Phitayakorn R. Adaptation to open-book online examination during the COVID-19 pandemic. *J Surg Educ*. 2020;78(3):737-739.
33. Alessio HM, Malay N, Maurer K, Bailer AJ, Rubin B. Interaction of proctoring and student major on online test performance. *The International Review of Research in Open and Distributed Learning*. 2018;19(5). https://doi.org/10.19173/irrodl.v19i5.3698
34. Daffin LW Jr, Jones AA. Comparing student performance on proctored and non-proctored exams in online psychology courses. *Online Learn*. 2018;22(1).
35. Stack S. The impact of exam environments on student test scores in online courses. *J Criminal Justice Educ*. 2015;26:273-282.
36. Rios J, Liu O. Online proctored versus unproctored low-stakes internet test administration: is there differential test-taking behavior and performance? *Am J Dist Educ*. 2017;31:226-241.
37. Mohanna K, Patel A. Overview of open book-open web exam over blackboard under e-learning system. In 2015 Fifth International Conference on e-Learning (econf)
38. Johanns B, Dinkens A, Moore J. A systematic review comparing open-book and closed-book examinations: evaluating effects on development of critical thinking skills. *Nurse Educ Pract*. 2017;27:89-94.
39. Serafini F. Three paradigms of assessment: measurement, procedure and enquiry. *The Reading Teacher*. 2000;54:384–393.

40. Appiah M, and van Tonder F. E-assessment in higher education: a review. *Int J Bus Manage Econ Res*. 2018;9:1454-1460.

41. Tertiary Education Quality and Standards Association. (2017) Guidance note: course design (including Learning Outcomes and Assessment). (TEQSA, ed) Vol. Version 1.3, Australian Government, Canberra.

42. Dennick R, Wilkinson S, Purcell N. Online eAssessment: AMEE guide no. 39. *Med Teach*. 2009;31:192-206.

43. Murdock M, Brenneman M. Designing tests from question pools with efficiency, reliability and integrity. *J Contemp Chiropr*. 2020;3(1):92-100.

44. Zaidi NB, Hwang C, Scott S, Stallard S, Purkiss J, Hortsch M. Climbing Bloom's taxonomy pyramid: lessons from a graduate histology course. *Anatom Sci Educ*. 2017;10:456-464.

45. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *BMC Med Educ*. 2007;7:.

46. Brady A-M. Assessment of learning with multiple-choice questions. *Nurse Educ Pract*. 2005;5:238-242.

47. Haladyna TM, Downing SM, Rodriguez MC. A review of multiple-choice item-writing guidelines for classroom assessment. *Appl Measur Educ*. 2002;15:309-333.

48. Allanson PE, Notar CE. Writing multiple choice items that are reliable and valid. *Am Int J Human Social Sci*. 2019;5:1-9.

49. Crowe A, Dirks C, Wenderoth MP. Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sci Educ*. 2008;7:368-381.

50. Clusky GR Jr, Ehlen CR, Raiborn MH. Thwarting online exam cheating without proctor supervision. *J Acad Business Ethics*. 2011;4:1-7.

51. Michael T, Williams M. Student equity: discouraging cheating in online courses. *Admin Issues J*. 2013;3:6.

52. Zaidi NLB, Grob KL, Monrad SM, et al. Pushing critical thinking skills with multiple-choice questions: does Bloom's taxonomy work? *Acad Med*. 2018;93:856-859.

53. Denny P, Manoharan S, Speidel U, Russello G, Chang A. On the fairness of multiple-variant multiple-choice examinations. In Hawthorne E, Pérez-Quiñones M, eds. *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. Minneapolis, MN, USA: Association for Computing Machinery; 2019:462-468.

54. Monrad SU, Bibler Zaidi NL, Grob KL, et al. What faculty write versus what students see? Perspectives on multiple-choice questions using Bloom's taxonomy. *Med Teach*. 2021; 43:575-582.