ARE ONLINE EXAMINATIONS A VIABLE ALTERNATIVE TO PAPER-BASED EXAMINATIONS FOR ASSESSMENT OF HUMAN PHYSIOLOGY?

Karen Mate[a], Jude Weidenhofer[a]

Presenting Author: Karen Mate (Karen.Mate@newcastle.edu.au)
[a]School of Biomedical Sciences and Pharmacy, University of Newcastle, Callaghan NSW 2308, Australia

KEYWORDS: assessment, online assessment, academic integrity, Bloom's taxonomy

ABSTRACT

There are practical and pedagogical reasons for the increasing role of online assessment in higher education. This study examines student performance on paper-based and online examinations, varying both examination settings and proportions of questions coded by a modified Bloom's taxonomy, to inform the effective and sustainable assessment of first year students in an introductory human physiology course. Student performance was analysed across three delivery formats of a mid-semester multiple choice assessment of the same concepts. Delivery formats were either i) invigilated paper examination with questions presented in random order across three versions of the paper, ii) online non-invigilated with answers in random order and questions presented individually in random order with no ability to backtrack or, iii) online non-invigilated with answers in random order and questions presented individually in random order and the ability to backtrack. Allowing students to backtrack appeared to improve student time-management, with more students completing all questions in the examination with these settings. Questions classified according to a modified Bloom's taxonomy showed student performance in lower-level Bloom's questions was significantly higher in online formats, especially when backtracking was allowed. Performance in higher level questions did not vary across formats. As such an 'open book' online assessment can provide similar rigor and discriminating power as an invigilated assessment if consideration is given to adjusting towards a higher proportion of questions assessing higher order learning.

Proceedings of the Australian Conference on Science and Mathematics Education, 29 September - 1 October 2021, pages 78-83, ISSN 2653-0481

BACKGROUND

Summative assessments play an important role in ensuring students have factual knowledge, technical proficiencies, communication, and higher order cognitive skills. In the context of human physiology, students studying medicine, nursing and the allied health professions must meet the requirements of accrediting bodies. These requirements include a means of demonstrating that students have met key learning outcomes and standards, with summative assessments often performing this role. There are both efficiency and pedagogical reasons for the introduction and increasing role of online assessment, both formative and summative, in higher education (Boitshwarelo, Reedy & Billany, 2017; Gipps, 2005; Pauli & Ferrell, 2020), however these are balanced by practical challenges and risks, especially for summative assessments.

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 towards learning goals).There is a strong perception amongst academics (Reedy, Pfitzner, Rook, & Ellis, 2021) however, and some evidence (Cerimagic & Hasan, 2019; Dawson, 2021; Reedy et al., 2021), 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. The prevalence of cheating, or willingness to cheat, on graded assessments amongst tertiary students has increased steadily over

a number of years, with estimates ranging from 9% to as high as 90% (Burgason, Sefiha, & Briggs 2019).

The sudden shift to online assessment necessitated by the COVID-19 pandemic meant that, in many cases, there was insufficient time to consider best practice in the adaption of assessment practices to online delivery. Whilst formal exams may revert to paper-based invigilated format, it is likely that many continuous assessments will remain online permanently, hence the need to provide valid and reliable measures of student learning, that are accessible, secure and ensure academic integrity.

AIMS

This study aimed to identify measures for effective and sustainable assessment of first year students in a large cohort introductory human physiology course. The study investigated student performance on multiple choice examinations delivered as either paper-based or online examinations. Specifically performance across the whole paper, as well as performance at each of three modified Bloom's taxonomy levels was investigated, with consideration to invigilation, timing and ability to backtrack across three versions of the paper.

DESIGN AND METHODS

SETTING

This study investigated how the format of a mid-semester examination altered student ability to demonstrate their knowledge and understanding of the same concepts. HUBS1403 Biomedical Science I, a year one first semester course at the University of Newcastle, is the first part of an intensive introductory human physiology course covering the chemical, cellular and tissue levels of organisation, genetics, fast and slow control mechanisms of the endocrine and nervous systems and finishes with muscle function. It is a core course for students from multiple degree programs including biomedical science, pharmacy, physiotherapy, speech pathology, nutrition and dietetics, and podiatry with an intake of 550-650 students each year.

TEST FORMAT

Assessment traditionally included an invigilated paper-based mid-semester examination consisting of 30 multiple choice questions (MCQs). This assessment was transitioned to an online format in 2020 and 2021. Several strategies were implemented to address academic integrity of assessment online, including adjustments to the test settings in the learning management system (Table 1) as well as requiring students to commence the exam within a 30 minute time window. Students were also required to review and agree to an academic integrity statement before starting the test in 2020.

Table 1: Format of undergraduate first year physiology mid-semester exam (consisting of 30 multiple choice questions) during the period 2019-2021.

	Mode	Invigilated	Time Allowed	Number of MCQs at Bloom's Taxonomy Level^			Online Test Settings		
				1	2	3	Presentation	Order†	Backtracking
2019	Paper	Yes	45 min	13	11	6			
2020	Online	No	45 min	7	11	12	One-at-a-time	Random	Not Allowed
2021*	Online	No	40 min	11	11	7	One-at-a-time	Random	Allowed

† Question and answer options

* Only included 29 questions due to a technical error

^ Modified Bloom's taxonomy such that Level 1 involve recall, Level 2 require an amount of interpretation and Level 3 involve application and/or analysis

TEST CONTENT

The proportion of questions that tested higher order cognition, and therefore not easily searched on the internet, was increased when the test first transitioned online in 2020. Questions were classified into three categories according to a modified Bloom's taxonomy scale (Krathwohl, 2002); Level 1 questions involved simple recall, Level 2 questions required some amount of interpretation and Level 3 questions involved application and or analysis of key concepts. Examples of MCQs classified at each of these levels are provided in Table 2. The number of Level 1 recall questions was reduced

from 13 to 7 in 2020, with a concomitant increase in the number of Level 3 application/analysis questions. As a first year undergraduate course, it would be unreasonable to exclude Level 1 questions from a mid-semester assessment entirely. Therefore a smaller proportion were retained and the allowed time to complete the test limited such that it would be challenging to look up answers to all these questions and still have time to complete the assessment. Differences in distribution of questions in 2020 and 2021 arose through changes in time allowed for the assessment from 45 to 40 minutes, with a reduction in the number of Level 3 questions required to compensate.

ANALYSIS

Overall student performance on the mid-semester examination across three years was compared using Welch's ANOVA and Games-Howell post hoc test as there were unequal variance and sample sizes. Five questions at each Bloom's taxonomy level were selected based on their similar style and content in the 2019, 2020 and 2021 examinations to calculate average difficulty (% answered correctly).

Table 2: Example of multiple choice questions classified as each of the three Bloom's taxonomy levels.

Bloom's Level	Question Stem	Answer Options (correct bolded)
1	Which body fluid compartment contains high levels of K+, large anions and proteins?	**A. Intracellular fluid** B. Interstitial fluid C. Plasma D. Both plasma and interstitial fluid
2	The tissue in the image is best described as: 	A. Skeletal muscle B. Loose connective tissue C. A tendon or ligament **D. Dense regular connective tissue**
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** B. 10 C. -35 D. -60

RESULTS & DISCUSSION

OVERALL STUDENT PERFORMANCE

The average marks for the 2019 paper-based and 2020 online examinations were not significantly different ($p = 0.3$), however the 2021 average mark was significantly higher (both $p<0.001$; Table 3). It was notable that the variation in marks was significantly greater in the 2019 paper-based examination than the online examinations. Although the average mark did not differ in 2019 and 2020, the percentage of students that passed the examination (that is, scored at least 15 out of 30) was 10% higher in 2020. The percentage of students that passed the examination increased a further 10% in 2021 (Table 3).

Table 3: Performance of first year physiology students in mid-semester examinations held in invigilated paper-based (2019) and non-invigilated online (2020, 2021) formats.

Year	Number of Students	Average Mark	Standard Deviation[#]	% of Students Passing
2019	572	19.1	5.7	76.5%
2020	581	19.3	4.6	85.1%
2021*	627	21.9[†]	4.4	96.0%

***Individual marks adjusted from score out of 29 to score out of 30 to calculate average mark.
[†] Significantly different (P<0.01)
[#] Variances of the three samples are unequal

STUDENT PERFORMANCE BY MODIFIED BLOOM'S LEVEL OF QUESTION

Student performance in Level 1 recall questions was most affected by the shift from an invigilated paper-based to a non-invigilated online examination format. The average difficulty (percentage of students that answered the question correctly) increased by approximately 10% with the shift to online in 2020 and by another 10% in 2021. This increase from 2020 to 2021 may be due to allowing students to backtrack or may also reflect the different style of questions from the staff who taught the course in 2021. With backtracking allowed, as in a paper-based examination, students were able to return to Level 1 questions that lend themselves to searching for answers either in their notes or online. In contrast there was very little change in the difficulty of Level 2 interpretation questions or Level 3 application/analysis questions across the three formats (Figure 1), highlighting the importance of careful question selection.

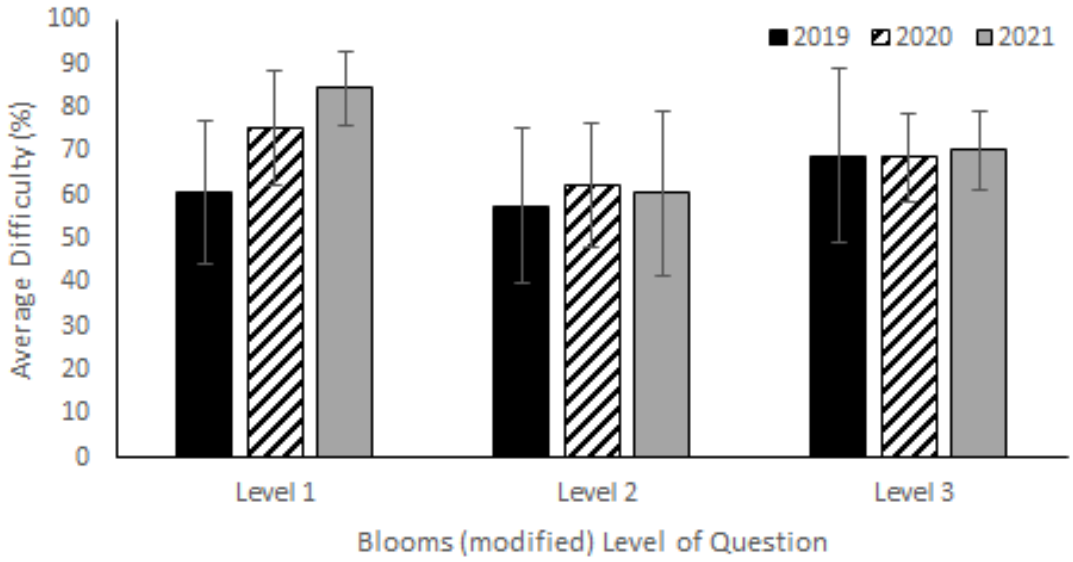

Figure 1: Student performance by questions categorised by modified Bloom's taxonomy across invigilated and non-invigilated examination formats

Level 1 questions were simple recall, Level 2 required some interpretation and Level 3 application and analysis. The examination in 2019 was invigilated paper based; 2020 & 2021 online non-invigilated questions presented one at a time in random order, 2021 allowed backtracking whereas 2020 did not.

TIME MANAGEMENT

From our experience, it is unusual for students to leave any questions unanswered in a MCQ examination as penalties are not applied for incorrect answers; only four students left a single question unanswered in the 2019 paper-based examination (Figure 2). In the 2020 online examination however, where backtracking was not permitted, there were 39 students that left at least one question unanswered and ten that left at least four questions unanswered (Figure 2). Many students (about 15%) provided unsolicited written or verbal feedback on difficulties managing their time under the settings employed in 2020; with the inability to backtrack or return to more complex questions a common concern. A number of students reported this 'on behalf of their friendship group' suggesting 15% is likely an under-representation of actual numbers. The random ordering of questions also

meant that some students could be presented with the most challenging questions first, affecting their confidence and time management. Allowing students to backtrack in 2021 reduced the number of students with unanswered questions when the test auto-submitted after the allocated time even though the total time allowed was reduced by five minutes.

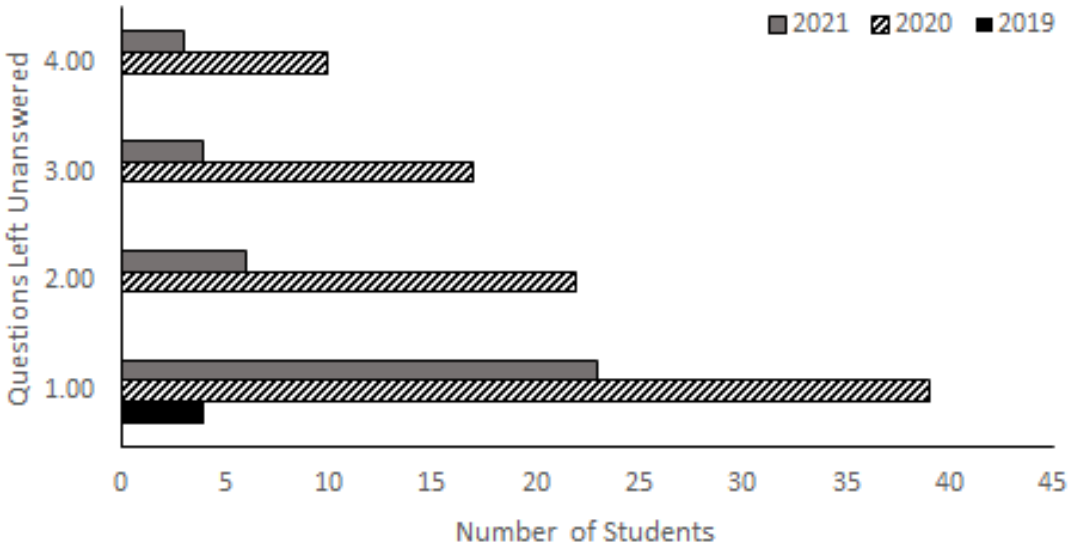

Figure 2: Effect of examination format on time management as determined by the number of students leaving questions unanswered.
The examination formats were: 2019, paper-based and 45 minutes duration; 2020, online 45 minutes duration with backtracking not allowed; 2021 online 40 minutes duration with backtracking allowed.

CONCLUSION

The analysis of student performance on examination questions of the same topics across three years of varied format examinations showed that an 'open book' online MCQ assessment can provide similar rigor and discriminating power as an invigilated assessment, if a few modifications are made. Most importantly, the increase in performance of 10-20% on L1 questions with no change to performance on L2 and L3 questions suggests the assessment should include a higher proportion of questions that assess higher order learning rather than recall. In this way the assessment is likely to be more reflective of student knowledge as opposed to confounded by an ability to 'look up' answers. The option to backtrack allows students to use the commonly used and effective strategy of returning to difficult questions after completing other questions, and therefore improved time management. However, the ability to backtrack may have presented increased opportunities to search for answers either in notes or online. As such the format of the examination must not be a 'standard' across the board approach, but rather requires nuancing based on the objectives of the assessment and course objectives being examined. Further, whilst not on their own likely to impact on cheating, the opportunities for online collaboration and sharing of answers can be limited when students all commence the exam at the same time and have only a limited but reasonable timeframe in which to complete the test with additional measures, such as randomisation of both questions and answers.

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.

REFERENCES

Boitshwarelo, B., Reedy, A. K., & Billany, T. (2017). Envisioning the use of online tests in assessing twenty-first century learning: a literature review. *Research and Practice in Technology Enhanced Learning 12*, 1-16.

Burgason, K. A., Sefiha, O., & Briggs, L. (2019). Cheating is in the eye of the beholder: An evolving understanding of academic misconduct. *Innovative Higher Education 44*, 203-218

Cerimagic, S., & Hasan, M.R. (2019). Online Exam Vigilantes at Australian Universities: Student Academic Fraudulence and the Role of Universities to Counteract. Universal Journal of Educational Research, 7, 929-936.

Dawson, P. (2021). Strategies for using online invigilated exams. Resource of the TEQSA Online learning good practice. Accessed August 13th 2021.

Gipps, C. V. (2005). What is the role for ICT-based assessment in universities? *Studies in Higher Education 30*, 171-180
Krathwohl, D.R. (2002) A revision of Blooms taxonomy: an overview. *Theory into Practice, 41*(4), 212-218
Pauli, M. & Ferrell, G. (2020). The future of assessment: five principles, five targets for 2025. Retrieved June 18, 2021, from https://www.jisc.ac.uk/reports/the-future-of-assessment
Reedy, A., Pfitzner, D., Rook, L., & Ellis L. (2021). Responding to the COVID-19 emergency: student and academic stagg perceptions of academic integrity in the transition to online exams at three Australian universities. *International Journal for Educational Integrity 17*, Article number 9
Reich, R.W., Milhano, C., & Valentine, D. (2018). *A comparison of proctored on ground exams vs unproctored online exams in undergraduate finance courses.* The Academy of Business Research International Conference, 2018, Boston, USA.