



Adding authenticity to controlled conditions assessment: introduction of an online, open book, essay based exam

Chris P. Moore 

Correspondence: Chris3.moore@uwe.ac.uk

Department of Applied Sciences,
University of the West of England,
Coldharbour lane, Bristol BS16 1QY, UK

Abstract

In the practice of designing examinations, where the emphasis is normally on the assessment of knowledge and understanding of content covered, the authenticity of that assessment model related to skills required for employability, as well as our growing access to instant information, is often missed. Some traditional methods of this form of assessment, particularly the paper based essay exam, have expectations of writing quality and structure as well as of knowledge and understanding, but without providing the tools with which to appropriately meet those expectations. To address this, a computer based, unseen, essay format exam was introduced to a final year module for applied science students, where access was permitted to specific online journal resources. Students wrote their submissions under controlled conditions in Microsoft Word and submitted it through the Virtual Learning Environment (VLE), Blackboard for grading. In addition to an improvement in average mark across the cohort compared to the previous year of paper based essay examinations on the same module, student feedback highlighted an improvement in their ability to organise and arrange their knowledge, and the usefulness of using research material to better evidence their own knowledge. Feedback, being completed online and returned to the students also provided greater understanding of grades achieved and of how to improve for future examinations, a unusual process for exams where graded papers are not normally returned.

Though upskilling of staff is likely required for more widespread use, and repeat cohorts will no doubt demonstrate further pros and cons of this form of assessment, the model presented here shows promise for assessment of student communication of depth and breadth of knowledge, and of demonstrating skills more relevant to future employment.

Introduction

Controlled condition assessment remains a key determinant and valuable metric of ownership of student knowledge and understanding. Coursework is a useful tool for assessing use of knowledge and its application, and of key skills be them written, communicative, or practical. But the absence of constraints such as time (before the approved deadline and assuming appropriately early engagement with the process), and having the ability to access and use information from virtually anywhere, does not guarantee proof a student who attains a high mark for a piece of coursework has sufficiently assimilated their learning material for effective use in employment.

Exams therefore, provide a model for evidencing that assimilation so that academics and, more importantly, the student can be confident they can draw on relevant knowledge when required. Or at least have developed the skill to draw on learned knowledge when necessary, even if that knowledge is ultimately outside the field of their programme of study. This knowledge may only be retained in the short term due to the anecdotal common practice of surface learning for the purpose of assessments (Godor, 2016), or ultimately forgotten once that academic year has passed once focus shifts according to new demands (Custers, 2010).

Traditional essay based exams however, though often believed to be the most effective measure of knowledge gain (Stanger-Hall, 2011), particularly in the final year of undergraduate education, fail to do something that coursework achieves very well; to simulate a real-world working environment. There are few instances in employment where, even to a deadline, an employee is tasked with sitting in a room alone with no access to the internet or other source of information beyond their own memory, and made to write a high quality account with only a pen and paper with which to record and properly organise their knowledge. Instead, in this digital age, the strength in writing effectively becomes not so much in the information, but in how it is communicated. Factual recall remains a crucial part of any career. But this can be assessed through alternative assessment methods such as MCQs, short answer questions, and Vivas or OSCEs (Objective Structured Clinical Examinations). The skill of writing comprehensively, persuasively, and clearly on the other hand, remains, at least outside of coursework, one assessed by essay questions or case studies, conducted under controlled conditions.

There are of course attempts to rectify this. Those of the open book exam (OBE), where traditionally the student is permitted to bring a (normally) fixed volume of prepared notes or similar material. This approach however, is usually done in tandem with the students either being informed of the question(s) beforehand so they know which notes to prepare. Or by providing a small bank of questions, only some of which will appear in the actual exam. Student approaches to these are somewhat different than to closed book exams (CBE) where often the assumption that access to prepared material in some way nullifies the need for extensive revision as in the case on CBEs. Nonetheless, when considering numerous factors that can impact exam performance such as preparation, duration or work when in the exam, and student perceptions before, during, and after, there appears to be no conclusive evidence as to which benefits student learning best (Durning, Dong, Ratcliffe, Schuwirth, Artino Jr, et al., 2016). Though Moore and Jensen (2007) did report that students who undertook OBEs, though scoring higher than those who took CBEs, displayed diminished academic behaviours in terms of class attendance and extra credit submission compared to their CBE counterparts.

However, even OBEs are missing one key element if compared to a real-life employment situation; digitisation. Though an OBE addresses the access to information problem by providing *some* material to use in order to expand on one's own knowledge, the pressure is still on the student to communicate that supplementary information effectively with only the medium of pen and paper with which to get that process right *the first time*. Again, in the working environment, an author of any document has the means to rearrange, edit, and reconsider.

It could be argued that use of an offline computer solves this problem, and it is indeed likely that this would have a positive impact on student anxiety if they knew they could rethink and reorganise their thoughts before submission (Martinez, Kock, & Cass, 2011). But the prepared notes still remain a finite source of information that students may rely too heavily on, to the detriment of their own revision based learning.

Lastly in all cases of paper based exams, the matter of useful feedback remains an issue. Though it is possible a student may wish to see their exam paper and discuss feedback with the marker, this is rarely the case due to so many student assignments now being submitted and marked online. Feedback on paper based assignments and examinations tends to be, with some exceptions, sparse and lacking in sufficient detail for the student to act on it and improve their performance for next time.

In order to address these issues, and to answer the question of whether it is possible to create an exam that closely mimics a working environment situation, yet remains under controlled conditions and robust as an assessment model, a cohort of final year undergraduates at Bristol UWE undertook a single unseen question, computer written and online open book exam with strategic limitations.

Methods

A cohort of 88 final year students enrolled on Biomedical, Biological, or Forensic Sciences at the University of the West of England (UWE) Bristol and studying the module *pathophysiology* undertook this new form of assessment. The assessment was conducted on campus, in PC labs, and invigilated following standard UWE controlled conditions assessment regulations. All students signed an attendance sheet and student ID cards were checked to confirm identity and eligibility to sit.

Format and expectations

The exam consisted of a single, unseen, essay style question to be attempted over a period of three hours; the time usually set aside for three essay questions in paper based format. This question was provided via Blackboard on time release to become available at the start of the exam. Students were told verbally and via guidance provided online within the exam area of the module on Blackboard that they must include a minimum of five references in their answer, properly cited and referenced, and that images were not permitted. Answers were written in Microsoft word, which was the only document permitted to be opened.

Constraints

Students were permitted to access the UWE library website and any journal websites that searches from the library linked to. Additional journal search resources were limited to Pubmed and Google Scholar. Students were not permitted to open any tabs beyond these resources. All content areas of the module Blackboard page were hidden for the duration of the exam, with the exception of the exam area itself, which in addition to the exam question contained direct weblinks to the permitted resources, guidance on how to complete the exam, and the assignment submission tool.

Submission

On completion of the essay, students were instructed to submit their saved Microsoft Word document via the assignment tool within Blackboard. In order to prevent and/or monitor access to materials or websites outside of those permitted, students took screenshots of their browser and download histories and uploaded them as image files alongside their essay submissions.

Marking

All submissions were automatically put through the online plagiarism checking tool, *Safe Assign*, then distributed randomly among the marking team via the UWE online marking tool within Blackboard. Markers were given clear instructions on the key elements expected to be present in submissions that appropriately addressed the question, and were provided with the faculty level marking criteria for final year assessments. To optimise standardisation, markers were instructed to assign final marks for each essay based on the number of criteria met within each grade boundary, modifying the final grade depending on numbers of criteria met across boundaries. Feedback was to be provided throughout each submission by way of comments using the review tool in Word.

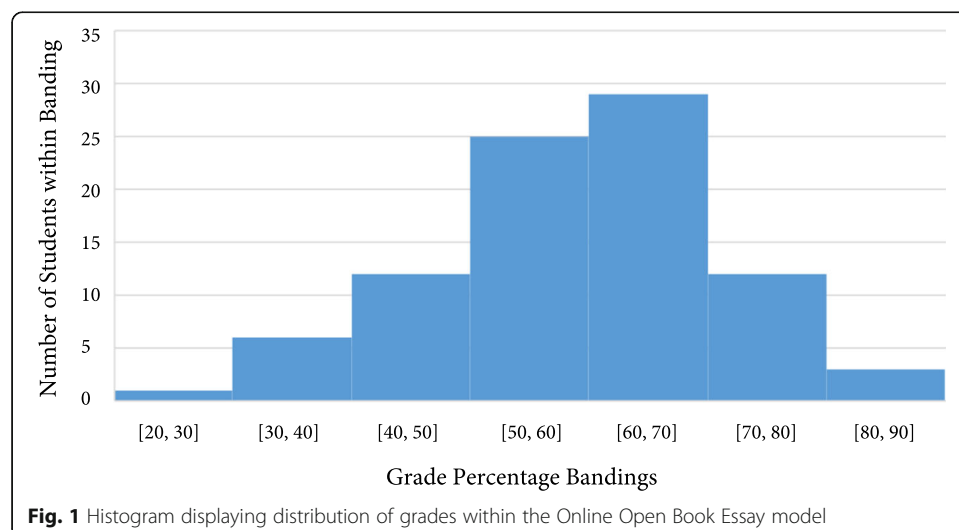
Results

Student grades for this exam were compared with those of the previous academic year where the format consisted of three essay questions from a choice of six, set over a three hour period and done in a traditional paper based setting.

The grade distribution depicted in Fig. 1 shows a very close binomial distribution in slight favour of upper tier grade boundaries. The average percentage mark for the 88 student cohort was 60.38% compared to a previous average for the 2016/17 paper with 61 students and an exam average of 52.8%.

There were several trends of note that related content to grade given:

Students who included significantly more references than the minimum required, on average received higher grades.



Students who used their references to evidence that which they had relayed from their own knowledge of the lecture material received higher grades than those who used extensive referencing to make up for a shortfall in knowledge that was clear to have been assimilated through engagement with the lecture material.

Critical evaluation of the question was more apparent in those who used the resources available to them to include evidence of new research in the area in which the question was based. Those with a notably weaker grasp of the core material showed a tendency to use evidence from the resources available to reference definitions and well accepted outcomes than to further and add validity to their own well-informed opinion as in those with a stronger understanding of the core material.

Discussion

Uptake of universities to new forms of computer based assessment can be slow due to the increased need for more resources and space, and the need for an acceleration in the digital agility of academic staff and invigilators (Boevé, Meijer, Albers, Beetsma, & Bosker, 2015). This reluctance of staff to learn new skills, or difficulty with learning those skills required, combined with infrastructure issues can lead to an exacerbation of the fear of digital based assessments among students. This is despite the reality of most forms of employment, particularly in the scientific sector, involving significant use of technology. In the case of the model of digital assessment tested here, several students in the cohort expressed some initial concern over the process. These concerns were, for the most part alleviated through implementation of a workshop and mock version of the exam six weeks prior to the actual exam. But this does raise the interesting point of, despite the perceived ease of such an assessment format, given the nature of today's students engaged with mobile devices and social media, the importance of the same reasonable adjustments being considered as with traditional forms of assessment. In short, it can be all too easy to presume student perceptions of new assessments relating to digital innovation will always be positive, simply based on their use of online social media platforms. But as Struyven, Dochy, and Janssens (2003) suggest, the idea of the student as an active partner in the development of new learning and assessment methods remains an important one.

There were a number of minor technical issues that arose during the course of the exam. Chief among these was a period after the exam ended when submissions were to be uploaded where approximately 75% of the PCs being used experienced a significant lag, delaying submission for many students. However, as this occurred after the assessment itself was over, and on discussion with several students afterwards, it is not believed this could in any way have impacted the work done during the assessment period, or affected results. A significant number of students also required repeated instruction on how to create screenshots and save them, which again was not an issue that arose during the exam itself and so was more an experiential concern than an assessment one.

Surprisingly, given the far longer timeframe to answer a single question than a standard exam, most students used all the time available to them, writing significantly more than was expected. It was believed that, despite having three times longer than traditional paper based essays exams, students would write only until they felt they had

answered all they could or otherwise experience exam fatigue. Total word count was therefore expected only to be around 500 words, plus references. Most students however, wrote in excess of 2000 words, plus references. Student feedback provided after the exam suggested the longer timeframe provided the opportunity to better evidence and edit their work, improving the readability and professional quality of their submissions significantly. This seems to disagree with observations made by Moge and Fluck (2015) who found in a comprehensive study of factors affecting student preferences to paper versus online exams that although students were aware of the importance of editing in structure, they seemed more concerned with the volume of text they could get down. Originally the longer duration of the exam was set to allow for variations in typing speed among the cohort, an impact on writing quality observed by Connelly, Dockrell, and Barnett (2005). This did not seem to be a necessary factor, though as it encouraged better engagement with the process it can only be seen as a positive and will likely be retained in future iterations of this assessment.

One feature of this type of assessment in particular, stemmed from a huge deviation in the way in which students receive their marks when undertaking traditional paper-based forms of assessment; that of feedback and feed forward. At this institution at least, students do not receive their graded exam papers back and therefore have little awareness beyond their mark (often an average of several questions attempted) as to how they did and how they might improve. Given the weighting and importance put upon controlled conditions assessments, this seems a strange practice to continue to engage with, particularly in the current climate of increasing student demand for higher quantity and quality of feedback. The model used in this paper, in part due to its place in the academic year of being sat after the first semester, provided the opportunity for students to receive feedback in both areas comparable to that which they should receive in any piece of coursework; feedback that may be used to understand how a grade decision was reached, and how strengths and weaknesses may be applied and improved in future pieces of work (irrespective of module or content) that require similar writing skills. In short, a *feed forward* mechanism of exam performance was possible for modules with exams at the end of the academic year. Though the format of these would of course be different, the lessons learned about (for example) answering a question properly over demonstrating how much one knew about the topic regardless of its relevance to the question being asked are adaptable to more traditional, paper based exams. But still beyond that possible with feedback from coursework that is produced in an entirely different manner and environment.

Comparison of results in this type of pedagogic study remains difficult. Using the results of assessments from previous cohorts of students does risk the impact of cohort variation and is not a true standardised experiment. In this instance in particular, the number of questions and the exclusion of choice greatly differentiates the format of the assessment when compared to the previous year, increasing the variables beyond those of the primary study of online open book versus paper based closed book. However, in the exam tested here the deliberate openness of the single question was designed to promote the inclusion of multiple topics from those studied on the module, simulating the coverage one might expect from three essays chosen from six, each focussed on (usually) a single topic from the module. The depth and breadth of learning material covered in this exam therefore, exceeds if not approximates that of a traditional paper

based exam. Further, by using a single, *unseen* question, students had no choice but to avoid “question spotting” where they may normally revise only those topics they planned on using in their exam. Instead revising the entirety of the module, which ultimately benefitted them by including multiple, and varied topics related to pathology. And on seeing the question set as the exam began, they became aware that only those submissions with well written depth *and* breadth would likely receive first class marks. This led to a richer reading and marking experience for the module team due to variety of student submission.

It must be considered however, that it is possible that the positive response of students to the assessment, and their extensive preparation beforehand may have been influenced by the fact that they had not experienced an assessment of this type before. Several studies have explored the nature of anxiety related to transitioning from paper based to online assessments (Stowell & Bennett, 2010; Schult & McIntosh, 2004), which the model tested here could certainly apply. It could be surmised therefore, that if this form of assessment became more commonplace across the programme, that student engagement with it before and during could wane in a similar manner to existing models.

Though further repeats of this process are sure to shed more light on the reproducibility of this experience when considering cohort variance. The engagement of the students and their performance as well as the clear benefits of more detailed and accessible feedback when compared to traditional forms of written exam feedback, suggest this approach is a positive one that maintains an academic robustness despite access to a wider body of information (though limited in scope to peer reviewed papers) than other models of open book exams. Further it could be argued that cohort variance is less of an impacting factor than one might imagine, provided we maintain a comparison of outcomes and experiences within the past five years. This is due to a fundamental change in students in terms of their approaches and attitudes to learning as well as metric and university strategies related to the student journey, not to mention advances in technology and technology awareness. This paradigm shift in higher education practice may raise questions regarding mapping conclusions related to student experiences and opinions of open book exams and digital assessments from much earlier work conducted in the late 1990's or even late 2000's in many cases. For example Spector (2000) noted that technology had yet to demonstrate any significant improvements in learning. While Shaffer and Resnick (1999) believed technology could be used to create authentic contexts for learning, and provide resources that give students opportunities in a number of areas. And McLoughlin and Luca (2001) ultimately came to the conclusion, on the basis of many works around the same time, that web based learning and assessment would continue to expand and that universities would have to offer more flexible approaches to both learning and assessment. The key to success here was believed to be in the design, the delivery, and the transparency of benchmarks to the students undertaking such activities.

The model presented here shows promise from an academic standpoint related to performance, realism, and authenticity, and was generally well received by the students. The most difficult barrier to break through when it comes to implementation however, appears to be hesitance to change by academics, or their perception of an inherent difficulty to learn and manipulate new technologies. It could be argued that some studies

that have concluded that online assessments are too problematic and do not increase student performance, such as Ricketts and Wilks (2010), may be approaching online assessment from the wrong angle. Modification of old or introduction of new assessments should not be about improving the performance of students taking them, but instead be about the authenticity to better prepare students for their world beyond university, thus encouraging better performance.

Availability of data and materials

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request (Please contact author for data requests).

Authors' contributions

CM conceived of, designed, coordinated, and implemented the study presented here. CM also authored the paper and carried out the data analysis. The author read and approved the final manuscript.

Competing interests

The author declares that he has no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 24 March 2018 Accepted: 2 May 2018

Published online: 03 July 2018

References

- Boevé, A. J., Meijer, R. R., Albers, C. J., Beetsma, Y., & Bosker, R. J. (2015). Introducing computer-based testing in high-stakes exams in higher education: Results of a field experiment. *PLoS One*, *10*(12), e0143616.
- Connolly, V., Dockrell, J. E., & Barnett, J. (2005). The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology*, *25*(1), 99–107.
- Custers, E. J. (2010). Long-term retention of basic science knowledge: A review study. *Advances in Health Sciences Education*, *15*(1), 109–128.
- Durning, S. J., Dong, T., Ratcliffe, T., Schuwirth, L., Artino Jr., A. R., Boulet, J. R., & Eva, K. (2016). Comparing open-book and closed-book examinations: A systematic review. *Academic Medicine*, *91*(4), 583–599.
- Godor, B. P. (2016). Moving beyond the deep and surface dichotomy; using Q methodology to explore students' approaches to studying. *Teaching in Higher Education*, *21*(2), 207–218.
- Martinez, C. T., Kock, N., & Cass, J. (2011). Pain and pleasure in short essay writing: Factors predicting university students' writing anxiety and writing self-efficacy. *Journal of Adolescent & Adult Literacy*, *54*(5), 351–360.
- McLoughlin, C., & Luca, J. (2001). Quality in online delivery: What does it mean for assessment in e-learning environments. In *A proceedings of the annual conference of the Australasian Society for Computers in learning in tertiary education (ASCLITE)*, (vol. 18).
- Mogey, N., & Fluck, A. (2015). Factors influencing student preference when comparing handwriting and typing for essay style examinations. *British Journal of Educational Technology*, *46*(4), 793–802.
- Moore, R., & Jensen, P. A. (2007). Do open-book exams impede long-term learning in introductory biology courses? *Journal of College Science Teaching; Washington Vol.*, *36*(7), 46–49.
- Ricketts, C., & Wilks, S. J. (2010). Improving student performance through computer-based assessment: Insights from recent research. *Assessment & evaluation in higher education*, *27*(5), 475–479.
- Schult, C. A., & McIntosh, J. L. (2004). Employing computer-administered exams in general psychology: Student anxiety and expectations. *Teaching of Psychology*, *31*, 209–211.
- Shaffer, D. W., & Resnick, M. (1999). "Thick" authenticity: New media and authentic learning. *Journal of Interactive Learning Research*, *10*(2), 195.
- Spector, M. (2000). Designing technology enhanced learning environments. In B. Abbey (Ed.), *Instructional and cognitive impacts of web-based education*, (pp. 241–261). Hershey: Idea Group Publishing.
- Stanger-Hall, K. F. (2011). Multiple-choice exams: An obstacle for higher-level thinking in introductory science classes. *CBE – Life Sciences Education*. Vol. *11*(3), 294–306.
- Stowell, J. R., & Bennett, D. (2010). Effects of online testing on student exam performance and test anxiety. *Journal of Educational Computing Research*, *42*(2), 161–171.
- Struyven, K., Dochy, F., & Janssens, S. (2003). Students' perceptions about new modes of assessment in higher education: A review. In *Optimising new modes of assessment: In search of qualities and standards*, (pp. 171–223). Netherlands: Springer.