Article

E-Examinations and the Student Experience Regarding Appropriateness of Assessment and Course Quality in Science and Medical Science

Ananthan Ambikairajah[1] and
Christopher C. Tisdell[2,3]

Abstract

The interest in, and use of, computers and software for assessment is reported to be increasingly popular via electronic examinations (e-exams). We deepen our understanding of the design, reception, and effectiveness of e-exams for history and philosophy of science modules, undertaken by first-year advanced science and medical science students at university. We employ a quasi-experimental research design approach to examine our implementation of e-exams on reported student satisfaction regarding the suitability of the information provided about the assessment requirements, appropriateness of the assessment methods, and overall quality of the associated courses. We report statistically significant increases in student satisfaction regarding the suitability and appropriateness of the assessment methods or requirements. The outcomes of this research highlight new avenues for educators to explore including (a) the innovative use of associated software (Maple TA[TM]) for

[1]Centre for Research on Ageing, Health and Wellbeing, Australian National University, Canberra, Australia
[2]School of Mathematics & Statistics, The University of New South Wales, Sydney, Australia
[3]Institute for Teaching and Learning Innovation, The University of Queensland, Brisbane, Australia

Corresponding Author:
Christopher C. Tisdell, School of Mathematics & Statistics, The University of New South Wales, Sydney, NSW 2052, Australia.
Email: cct@unsw.edu.au

e-exams and (b) the implications that e-exams can have on the student experience in the context of medium-stakes testing.

Keywords
e-examinations, student perspectives, digital education, assessment, survey

Introduction

There is increasing governmental support to embed information and communications technology (ICT) into the classroom to enhance the student experience within Australia. This is evident through the digital education revolution, an Australian Government-funded educational reform program, which was launched in 2008 with the aim of increasing ICT proficiency for teachers and students by providing laptops to all public high schools. Furthermore, ICT has been integrated within the learning and teaching objectives of core educational documents and policies such as the Australian Professional Standards and the Australian Curriculum (Australian Curriculum Assessment and Reporting Authority, 2014; Board of Studies Teaching and Educational Standards NSW, 2014). The importance of ICT, specifically electronic examinations (e-exams), has been further emphasized as a priority by the Australian Government, with national standardized assessment tasks shifting from pen and paper to an e-examination format. For example, since 2011, the Science Validation of Assessment For Learning and Individual Development test, which enables schools to map their students' progress in the science key learning areas, was reengineered to be an entirely computerized exam. The evolution of Australian learning and teaching practices emphasizes the increasing need to incorporate ICT centered examinations to meet the demands and technological learning requirements of 21-st century students.

At the tertiary level, the interest in, and use of, computers and software for assessment is reported to be increasingly popular (Hillier, 2014). Sindre and Vengendla (2015) predict a large-scale shift toward electronic examinations during the next 5 to 10 years.

According to Dawson's (2016) review of the literature, e-exams offer a number of advantages over paper-based examinations. For example, universities may appreciate the identified efficiencies in conducting electronic assessment delivery and feedback, including eliminating the process of handing out and then collecting paper scripts when invigilating an exam and the potential of automated marking.

Students may recognize the advantages in the e-exam process, such as the benefits of typing over writing. For instance, it has been shown that many

students prefer to type in exams (Mogey, Cowan, Paterson, & Purcell, 2012) and that typing has been correlated with clearer sentence structure in mock examinations (Mogey & Hartley, 2013). Furthermore, Sorensen's (2013) study demonstrated that students enjoy the flexibility of time and place offered by formative e-exams and felt that such assessment practices aligned well with modern e-learning approaches and added value to their learning.

Exam markers and assessors also have reason to prefer e-exams too, as there is no need to decipher messy handwriting and no pieces of paper to keep track of.

It is important to note, however, that there are also disadvantages that have been associated with e-exams. For example, students with sub-par typing ability may be disadvantaged, due to a reduced ability to review the whole test (Noubandegani, 2012). This issue is an inherent limitation with e-exams, although we would like to highlight that a pen and paper test can pose the same issues for students who struggle with writing.

Wibowo, Grandhi, Chugh, and Sawir (2016) reported students and staff being optimistic about e-exams, subject to improvements.

The security risk associated with e-exams is a core issue (von Grünigen et al., 2018) which cannot be overlooked. Hillier and Fluck (2013) call for more research into secure, reliable digital systems and procedures that offer a graduated transition pathway from pen to keyboard.

Ultimately, if e-exams are going to become an integral part of university assessment procedures, then we argue that there needs to be further research undertaken and evidence produced to indicate that e-exams can form a workable solution for students, teachers, and administrators.

In this article, we respond to the aforementioned needs by deepening our understanding of the design, reception, and effectiveness of e-examinations within the tertiary sector, for history and philosophy of science modules within first-year advanced science and medical science courses. In particular, our research questions are centered around the effects of e-exams on student satisfaction regarding the assessment methods, the associated information provided, and the quality of the student experience regarding these courses.

Our work differs from previous studies in the following ways: Our e-exams are neither high stakes nor low stakes. Instead, we classify them as medium stakes, where there are *some medium, non-trivial* consequences for the students sitting them. This contrasts with the vast majority of e-exams research which lie at the extremes of the spectrum, that is, low-stakes (no consequences for the student) or high-stakes examinations (important consequences for the student; Shepherd & Godwin, 2004). The impact of e-exams within a medium-stakes environment appears to be overlooked and underresearched in the education literature.

Our e-exams do not involve the built-in assessment features of Blackboard, Moodle, or QuestionMark Perception (Hillier, 2014). Rather, we employ

Maple TA^{TM} as the software for our assessment tool. Furthermore, we have applied Maple TA^{TM} to a nonstandard educational environment, through the use of this traditional mathematical software platform within nonmathematical courses.

Our e-exams do not involve distance education, or assessments which include a "bring your own device" policy (Hiller, 2014). Instead, our assessment procedure involved students completing the e-exams within a computer lab, using machines with a standard operating environment (SOE) and were invigilated face-to-face.

Our research methods differ from previous work, too. We take a quasi-experimental research design approach through the implementation of an intervention. Furthermore, we address a limitation within the experimental design of many e-exam studies by including a control group.

Our idea to survey students after the intervention aligns with the work of Lim, Ong, Wilder-Smith, and Seet (2006), Dermo (2012), and Sorenson (2013), and it thus presents an alternative from Hillier's (2014) preconception surveys. One potential risk with preconception surveys is that people are prepared to give opinions about things which they have no knowledge of (Coe, Waring, Hedges, & Arthur, 2017). We think it is highly appropriate to do a survey after students have experienced the intervention and when the students are more knowledgeable about various aspects of e-examinations.

Our instruments involve questionnaires embedded within course surveys. The results of student and course evaluations contribute to promotion, tenure and merit pay decisions, and consequently, generate controversy among faculty (Dommeyer, Baum, Hanna, & Chapman, 2004). In particular, some teachers are concerned whether changes to assessment in their courses might negatively impact student perceptions and course ratings. Thus, we feel it is of significant importance to see what effect, if any, the introduction of e-exams had on student satisfaction of course quality and the student experience within the associated courses.

Research Questions

Our Introduction section naturally leads us to the following research questions concerning student perceptions of e-exams and the courses in which they were used:

- Did students feel they were provided with clear information about the assessment requirements?
- Did students feel that the assessment methods were appropriate?
- Did the introduction of e-examinations have any effect on the overall student satisfaction regarding the quality of the associated courses?

Methods

In late 2015, a research project was funded by the university to investigate e-exams. This action aligned with calls from Cook and Jenkins (2010) for the need for organizations to support e-assessment and for a project to be put in place to introduce e-exams (Crisp, 2011).

The team listed on the project included a rich diversity of people. For example, our team included researchers, teachers, course coordinators, educational leaders, tutors, and academics from within, and external to, the faculty of science (see Acknowledgments section). This brought together a wealth of experience and a diverse set of expertise and perspectives, aligning with the suggestion of Crisp (2011) that the associated work is done by teams instead of individual teachers only.

Ethics approval for the research project was sought from, and granted by, the university, with the research plan adhering to university and national guidelines.

The project team clarified the desirable characteristics of possible assessment software tools to be used for the e-exams, which were as follows:

- Low barriers to implementation;
- Able to handle open-ended, short paragraph responses.

Through discussions, the team agreed on the software choice of Maple TA^{TM}. Maple TA^{TM} is an online assessment system, primarily aimed at science, technology, engineering and mathematics courses to "truly assess student understanding of math-based concepts" (Maple^{TM}, 2018). However, our use of Maple TA^{TM} deviates considerably from this aim through our application of it to courses with history and philosophy of science and medical science curricula.

One reason for choosing Maple TA^{TM} was due to the university already possessing a license to use the software on SOE machines across the campus. This meant that there was no additional subscription or licensing fees to be paid. In addition, at least one of the project team had used the software before, meaning that it was not completely foreign to the entire team and did not need to be learnt by all. Thus, we agreed that in this situation, the barriers to implementation were low.

Crucially, Maple TA^{TM} also supports open-ended, short paragraph responses, although this is perhaps not widely known or used. Furthermore, Maple TA^{TM} includes a text editor, meaning that students can easily edit or highlight parts of their responses during the e-exams.

We argue that using Maple TA^{TM} for nonmathematical courses with content like history and philosophy of science and medical science is highly innovative and opens alternative perspectives from outside the traditional boundaries of this software. This embodies a more radical approach to education, such as creating pedagogical *heterotopias* (Tisdell, 2018a) by opening up new and alternative ways of learning and teaching (Tisdell, 2017, 2018b) through assessment.

We describe the magnitude of stakes for our e-exams as medium, with each e-exam counting 20% toward the final grade, and each course having two e-exams per semester. This contrasts with the vast majority of e-exams research in low-stakes (no consequences for the student) or high-stakes examinations (important consequences for the student). The medium-stakes area appears to be underresearched regarding e-exams. Our intended message to students was to convey that "this assessment matters" but it would not necessarily be terminal to a student's final grade if something went wrong.

We did not give students a choice between electronic or paper formats for their assessment. This was partly to ensure uniformity, but we recognized that this could be seen as being somewhat inflexible. To build student awareness of avenues for examination adjustments or special consideration (in case they felt like they needed this), we communicated these options to students through Moodle, in lectures, and in the course information pack.

We communicated the details of the e-exams (format, login procedure, etc.) to students in our courses using multiple methods. This included placing information and guidelines within the course information booklets, online via Moodle announcements, and face-to-face within lectures. An important step of this process was giving students early access to a *mock* e-exam on the Maple TA^{TM} platform a few weeks before the *real* e-exams. This was designed to ensure students had opportunities to investigate the Maple TA^{TM} environment and to explore the text editor before they had their e-exams. As a contingency, in each week of the scheduled e-exams, we timetabled an extra slot at the end of the week in case there was an emergency with any of the earlier e-exams (e.g., power outage, software failure, fire evacuation, university strike, etc.), leaving a time and space to rerun them.

Our approach to delivering the e-exam also involved making decisions around academic integrity and hardware. We chose to have the e-exams on-campus in a large computer lab with SOE machines that we could control. This aligns with Dawson's theory (2016) of e-exams that computer software security is dependent on hardware security.

Our computer lab spaces were shared between schools across the university to optimize its use throughout the semesters. This is in contrast to Hillier's (2014) warning of computer labs "laying empty" for most of the year.

Our e-exams were invigilated face-to-face. Students had to choose from a list of suitable times to take the e-exam and, when they arrived, they could login to their assigned computer in the lab via secure password. Their identity was verified by checking the photo on student identification cards against the student taking the e-exam. This goes further than, say, just the traditional password-based approach of authentication, because it has been deemed as inadequate because passwords can be shared easily (Gao, 2012).

The questions in the e-exam were open ended and invited students to write a few sentences to respond. For an example, see Figure 1. Sometimes, the e-exams
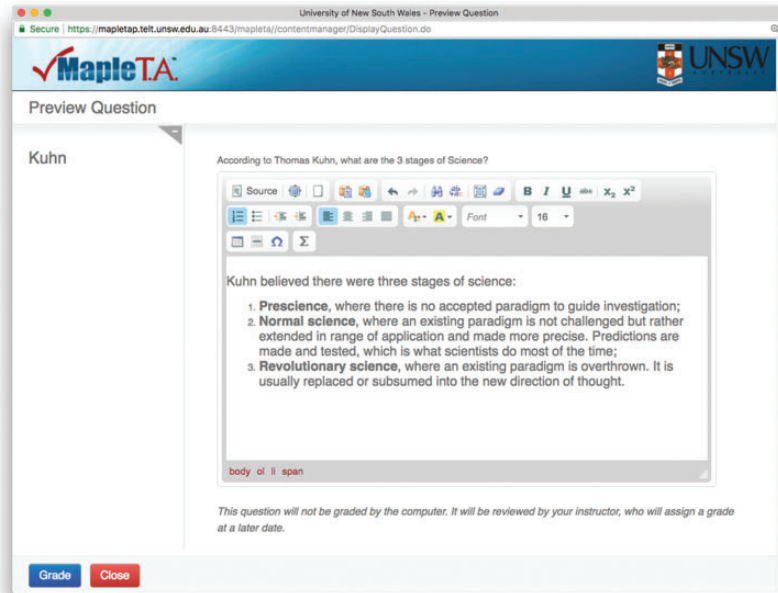
Figure 1. Example of Maple TA[TM] interface used for an open-ended question.

employed a combination of *factual* and *interpretive* questions. For example, the first part of a question might be factual and the second part interpretive.

Due to the free, open-ended nature of the questions and responses, the e-exams could not be automatically marked. Staff first had to read the responses and then mark them on their computer, or they could print out the scripts and provide comments and other feedback directly onto the article.

Methodological Position

As with most quasi-experimental designs (Morgan, 1997), our assumptions underpinning this work involve a realist ontology and a fallibilistic epistemology. Thus, our ideas draw on postpositivism.

Evaluation Overview

To evaluate the impact of our e-examinations, we designed and developed a quasi-experiment. Quasi-experiments share a common aim with all other experiments, namely, "to test descriptive hypotheses about manipulable causes, as well as many structural details, such as the frequent presence of control groups and pretest measures, to support a counterfactual inference about

what would have happened in the absence of treatment" (Shadish, Cook, & Campbell, 2001, p. 14).

Quasi-experimental evaluation design includes a control group and is preferred over nonexperimental approaches such as before-and-after design, for example, due to the vulnerability of a before-and-after design to internal validity threats (Robson, 2001).

Quasi-experiments provide a desirable alternative to traditional experiments when randomization is impractical or unethical. Because randomization is absent, this approach "provides a limited counterfactual which can infer limited causation" (Coe et al., 2017, p. 146). For example, it does not control for selection bias. Despite this, we can eliminate some threats to internal validity through the addition of certain elements to the basic quasi-experimental design, namely, a time series, which involves taking more measurements. The aims included: first, to establish a baseline trend; second, to make a comparison with the intervention phase.

A maturation threat is eliminated if we observe an abrupt change between the baseline phase and the intervention stage. Regression-to-the-mean or testing effects are reduced as possible threats if we can see that results are low before the intervention and repeatedly high afterwards. The threat of a history effect is somewhat reduced because the suitable interval for a coincidental event is narrowed by taking measurements more frequently.

Our experimental design involves two different stages (intervention and nonintervention) comprising two pairs of distinct phases:

- a baseline phase (2015);
- an intervention phase (2016–2017).

The timeline for these phases is summarized in Table 1.

As shown in Table 1, we ran quasi-experiments for the courses: Course 1 and Course 2.

In-line with the postpositivistic paradigm (Trochim, 2006), we embed multiple measures and observations into our approach and employ triangulation

Table 1. Experimental Design of the Study.

	S1 (2015)	S2 (2015)	S1 (2016)	S2 (2016)	S1 (2017)	S2 (2017)
Course number	1	2	1	2	1	2
Phase	O	O	X	X	X	X

Note. S1 = Semester 1; S2 = Semester 2; O = baseline, that is, no intervention; X = intervention, that is, the use of e-examinations.

across these sources. Furthermore, we shall use appropriate statistical techniques in the analysis and comparison of our data.

Our evaluation overview is summarized as follows in Table 2: Surveys for data collection serve important purposes in educational research (Berends, 2006). In Wang (2006, p. 36), we see reference to the term *survey* as "an instrument to collect data that describes one or more characteristics of a specific population."

With significant growth over the past 50 years, survey methods now form an important and accepted way of doing research in the social sciences. Both cost effective and time efficient, this method of research provides insight into the attitudes, thoughts, and opinions of populations.

Ideal for use in education, survey research is used to gather information about population groups to "learn about their characteristics, opinions, attitudes, or previous experiences" (Wang, 2009, p. 128).

There has been much debate regarding appropriate tests for ordinal data, particularly for Likert scales used in surveys. Traditionally, nonparametric tests (such as Mann–Whitney's *U* test) are considered appropriate; however, there is growing use of parametric tests (such as Student's *t* test) in the literature if sufficient sample sizes are available. The bottom line is that it is possible to employ both kinds of tests (Sullivan & Artino, 2013). Indeed, we include both of the aforementioned tests for completeness in our analysis.

Groups of Interest

The groups of interest in this research study are identified in Table 3.

Both Course 1 and Course 2 are taught at The University of New South Wales, which is a large, research intensive university located within Sydney, Australia.

Course 1 is an introductory medical science course with approximately 170 students and features modules on history and philosophy of medical science. Course 2 is a smaller introductory science course with modules on history and philosophy of science and approximately 70 students. The vast majority of students within these courses are in their first year of university studies.

Table 2. Evaluation Overview.

Evaluation approach	Timing or frequency or sample	Evaluation focus
Student attitude data	Quantitative attitudinal or perception data collected from students via course survey. 6-point Likertscale is employed, and mean values are calculated. Comparison between groups via Mann–Whitney *U* test and Student's *t* test at 95% confidence levels.	Impact on students' attitudes toward their learning, satisfaction, and quality

Table 3. Groups of Interest.

Group	Details
Target	Undergraduate students in science and medical science classes
Sample	Students in Courses 1 and 2 in the years of intervention
Comparison	Students in Courses 1 and 2 in the year of nonintervention

The courses run in separate semesters and are incompatible, in the sense that a student would not be allowed to complete both courses.

These two courses service students who are in the programs: Bachelor of Medical Science, Bachelor of Advanced Mathematics, and Bachelor of Advanced Science.

Randomization of students for this study was not possible due to economic and timetabling constraints. For example, there was a single lecture stream with all students enrolled within it. To set up a second stream was not possible due to the extra cost and resources required. This would have meant the strong possibility of timetable clashes for the students with other courses. Thus, random assignment of students was impractical in this case and meant that our situation was well suited to the quasi-experimental approach.

Results and Discussion

All statistical analyses were conducted using the open source software, R (version 3.3.3), running in RStudio (version 1.0.143). In each course where we implemented e-exams, students were asked to respond to three statements regarding their opinion on quality, communication, and appropriateness of assessment. The responses were assigned Likert values on a 6-point scale: $1 = Strongly\ Disagree$, $2 = Disagree$, $3 = Mildly\ Disagree$, $4 = Mildly\ Agree$, $5 = Agree$, and $6 = Strongly\ Agree$.

Let us report and examine the results from our surveys. Figures are rounded to four decimal places.

First Intervention Round (Course 1)

Mann–Whitney's U test and Student's t test were employed to compare student perception scores regarding the appropriateness of assessment tasks. There was a significant difference between the scores from 2015 (nonintervention) and 2016 and 2017 (intervention) regarding the appropriateness of assessment methods and tasks. See Table 4 and Figure 2 for more details.
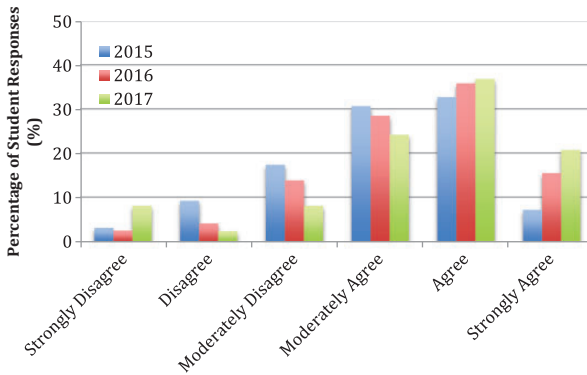
Two-independent sample Mann–Whitney U test and Student's t test were conducted to compare student perception scores regarding the clarity of the assessment requirements. There was a significant difference between the scores

Table 4. The Assessment Methods and Tasks in This Course Were Appropriate.

	n	Mann–Whitney U test	Student's t test
SI (2015)	98		
SI (2016)	123	**.0282**	**.0290**
SI (2017)[a]	87	**.0080**	**.0425**

Note. n = number of responses; SI = Semester 1. Bolded estimates indicate significance at the *p* <.05 level.
[a]2017 Question was amended to: "The assessment tasks were appropriate."

Figure 2. The assessment methods and tasks in this course were appropriate (Course 1).

from 2015 (nonintervention) and 2016 (intervention) regarding the clarity of assessment requirements. See Figure 3 and Table 5 for more details.

Mann–Whitney's *U* test and Student's *t* test were employed to compare student perception scores regarding the overall course satisfaction. There was no significant difference between the scores from 2015 (nonintervention) and 2016 and 2017 (intervention) regarding overall course satisfaction. See Table 6 and Figure 4 for more details.

Second Intervention Round (Course 2)

Two-independent sample Mann–Whitney *U* test and Student's *t* test were conducted to compare student perception scores regarding the appropriateness of assessment tasks. There was a significant difference between the scores from 2015 (nonintervention) and 2016 and 2017 (intervention) regarding the
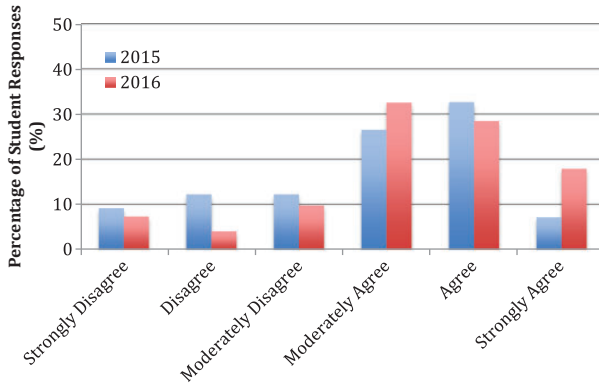
Figure 3. I was provided with clear information about the assessment requirements (Course 1).

Table 5. I Was Provided With Clear Information About the Assessment Requirements.

	n	Mann–Whitney U test	Student's t test
S1 (2015)	98		
S1 (2016)	123	**.0369**	**.0300**
S1 (2017)[a]	NA	NA	NA

Note. n = number of responses; NA = not applicable; S1 = Semester 1. Bolded estimates indicate significance at the $p < .05$ level.
[a]2017 Question not reasked.

Table 6. Overall, I Was Satisfied With the Quality of the Course.

	n	Mann–Whitney U test	Student's t test
S1 (2015)	97		
S1 (2016)	123	.7599	.4772
S1 (2017)	87	.3939	.4205

Note. n = number of responses; S1 = Semester 1. Bolded estimates indicate significance at the $p < .05$ level.

appropriateness of assessment methods and tasks. See Figure 5 and Table 7 for more details.

Mann–Whitney's *U* test and Student's *t* test were employed to compare student perception scores regarding the clarity of the assessment requirements.
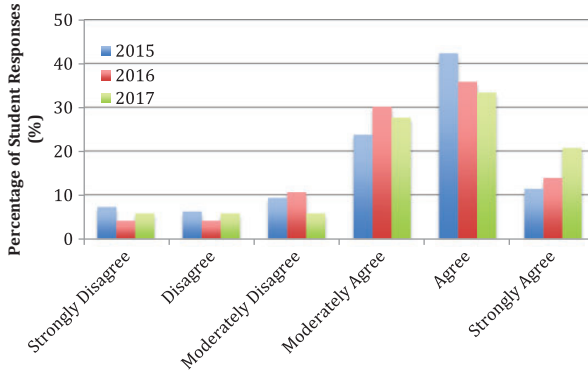
Figure 4. Overall, I was satisfied with the quality of the course (Course 1).
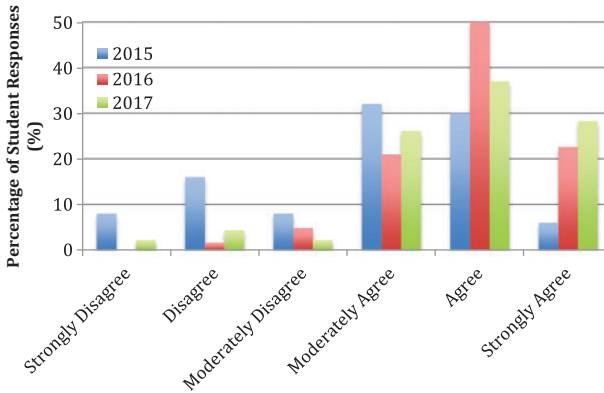

Figure 5. The assessment methods and tasks in this course were appropriate (Course 2).

Table 7. The Assessment Methods and Tasks in This Course Were Appropriate.

	n	Mann–Whitney U test	Student's t test
S1 (2015)	50		
S1 (2016)	62	**<.0001**	**<.0001**
S1 (2017)[a]	46	**.0003**	**.0003**

Note. n = number of responses; S1 = Semester 1. Bolded estimates indicate significance at the p <.05 level.
[a]2017 Question was amended to: "The assessment tasks were appropriate."
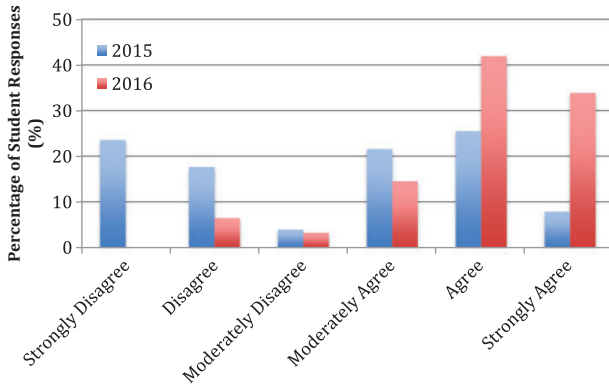
Figure 6. I was provided with clear information about the assessment requirements (Course 2).

Table 8. I Was Provided With Clear Information About the Assessment Requirements.

	n	Mann–Whitney U test	Student's t test
S1 (2015)	51		
S1 (2016)	62	**<.0001**	**<.0001**
S1 (2017)[a]	NA	NA	NA

Note. n = number of responses; NA = not applicable; S1 = Semester 1. Bolded estimates indicate significance at the $p < .05$ level.
[a]2017 Question not reasked.

There was a significant difference between the scores from 2015 (nonintervention) and 2016 (intervention) regarding the clarity of assessment requirements. See Figure 6 and Table 8 for more details.

Mann–Whitney's U test and Student's t test were employed to compare student perception scores regarding the overall course satisfaction. There was a significant difference between the scores from 2015 (nonintervention) and 2016 and 2017 (intervention) regarding overall course satisfaction. See Table 9 and Figure 7 for more details.

Benefits and Limitations

Let us discuss the limitations of this study by considering several threats to validity.

We acknowledge two main threats: an extreme or unusually low baseline measurement (regression to the mean) and nonrandomization of participants.

Table 9. Overall, I Was Satisfied With the Quality of the Course.

	n	Mann–Whitney U test	Student's t test
S1 (2015)	50		
S1 (2016)	62	**<.0001**	**<.0001**
S1 (2017)	46	.0049	.0148

Note. n = number of responses; S1 = Semester 1. Bolded estimates indicate significance at the *p* <.05 level.
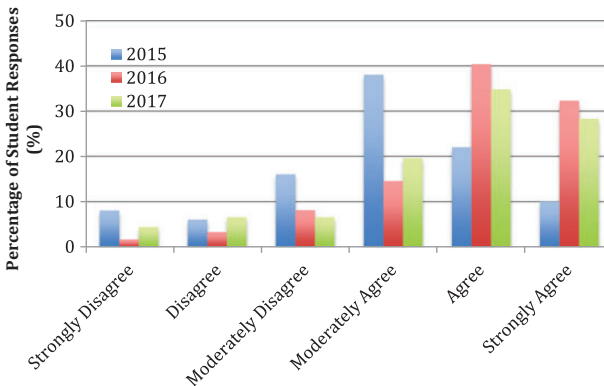
Figure 7. Overall, I was satisfied with the quality of the course (Course 2).

We cannot entirely rule out a low initial baseline effect having some influence on our findings. If we had several more years' worth of data before the intervention, then we could have reduced the threat of regression to the mean, as we would have a better idea of what the baseline results look like. However, there was no data available before 2015.

We also cannot absolutely rule out the possibility of selection bias having some influence during our phases. For example, one year's students may have been more open to the idea of e-exams than other years. As we explained in previous sections, randomization was impractical due to economic and timetable constraints.

One may wonder if there is also substantial qualitative data from students to further strengthen our position. However, this is not the case. Although we had healthy response rates to our surveys, only a tiny fraction of participants made additional (open ended) comments at the end of the survey. We thus do not report these due to the very small volume of returns.

In addition, we note that e-exams were not the only format of assessment within these courses. However, we argue that the instigation of e-exams was the

only change in assessment between the baseline and intervention years. Thus, we are confident that our results are capturing this change and that the data are relevant to the instigation.

Finally, we acknowledge that there remain important and challenging questions regarding e-exams. This is especially pertinent in the sciences and engineering, where complex formulae and technical diagrams feature within student-working in examinations. How to transfer this from handwriting to electronic formats remains open for investigation.

Conclusion

In this article, we responded to the need for research to be undertaken and evidence produced to indicate if and how e-exams can form a workable solution for students, teachers, and administrators. We deepened current understanding of the design, reception, and effectiveness of medium-stakes e-examinations concerning history and philosophy of science within first-year science and medical science courses at university. In particular, our research questions were centered around the effects of e-exams on student satisfaction regarding the assessment methods, the associated information provided, and the quality of the student experience regarding these courses.

Through quasi-experimental research design involving multiple e-exams over various courses throughout several years, our research found effects of significant increases in student satisfaction regarding:

- the suitability of the information provided about the assessment requirements;
- the appropriateness of the assessment methods; and
- and the overall quality of the associated courses.

The outcomes of this research open up new avenues for educators to explore, including how Maple TA^{TM} can be used in novel and different ways that are beyond the traditional mathematical testing format. We also call on researchers to widen their studies to medium-stakes examinations, rather than just concentrating on low- or high-stakes assessments.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

Australian Curriculum Assessment and Reporting Authority. (2014). *Foundation to Year 10 Australian curriculum.* Retrieved from http://www.syllabus.bostes.nsw.edu.au

Berends, M. (2006). Survey methods in educational research. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 623–640). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Board of Studies Teaching and Educational Standards NSW. (2014). *Australian professional standards for teachers.* Retrieved from http://www.aitsl.edu.au/australian-professional-standards-for-teachers/standards/list

Coe, R., Waring, M., Hedges, L. V., & Arthur, J. (2017). *Research methods & methodologies in education.* Los Angeles, CA: SAGE.

Cook, J., & Jenkins, V. (2010). *Getting started with e-assessment.* Bath, England: University of Bath.

Crisp, G. (2011). *Teacher's handbook on e-assessment, Transforming assessment—An ALTC fellowship activity.* Retrieved from http://transformingassessment.com/sites/default/files/files/Handbook_for_teachers.pdf

Dawson, P. (2016). Five ways to hack and cheat with bring-your-own-device electronic examinations. *British Journal of Educational Technology*, 47(4), 592–600. doi:10.1111/bjet.12246

Dermo, J. (2012). *Developing spaces for e-assessment: Cases from the University of Bradford. Transforming assessment webinar.* Retrieved from http://transformingassessment.com/events_3_october_2012.php

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, 29(5), 611–623.

Gao, Q. (2012). Biometric authentication to prevent e-cheating. *International Journal of Instructional Technology and Distance Learning*, 9(2), 15–20.

Hillier, M. (2014). The very idea of e-exams: Student (pre)conceptions. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Proceedings of ascilite on rhetoric and reality: Critical perspectives on educational technology* (pp. 77–88). Australasian Society for Computers in Learning in Tertiary Education: Tugun, Queensland.

Hillier, M. & Fluck, A. (2013). Arguing again for e-exams in high stakes examinations. In H. Carter, M. Gosper, & J. Hedberg (Eds.), *Proceedings of ascilite on electric dreams* (pp. 385–396). Australasian Society for Computers in Learning in Tertiary Education: Tungun, Queensland.

Lim, E. C. H., Ong, B. K. C., Wilder-Smith, E. P. V., & Seet, R. C. S. (2006). Computer-based versus pen-and-paper testing: Students' perception. *Annals of the Academy of Medicine*, 35(9), 599–603.

Maple[TM]. (2018). *Online assessment system for STEM courses.* Retrieved from https://www.maplesoft.com/products/mapleta/

Mogey, N., Cowan, J., Paterson, J., & Purcell, M. (2012). Students' choices between typing and hand-writing in examinations. *Active Learning in Higher Education*, *13*(2), 117–128. doi:10.1177/1469787412441297

Mogey, N., & Hartley, J. (2013). To write or to type? The effects of handwriting and word-processing on the written style of examination essays. *Innovations in Education and Teaching International*, *50*(1), 85–93. doi:10.1080/14703297.2012.748334

Morgan, G. (1997). *Beyond method: Strategies for social research*. Newbury Park, CA: Sage.

Noubandegani, P. A. (2012). Students' perceptions of computerized TOEFL test. *Language Testing in Asia*, *2*(2), 73. doi:10.1186/2229-0443-2-2-73

Robson, L. (2001). *Guide to evaluating the effectiveness of strategies for preventing work injuries: How to show whether a safety intervention really works*. Cincinnati, OH: DHHS.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.

Shepherd, E., & Godwin, J. (2004). *Assessments through the learning process. Questionmark white paper*. Retrieved from http://link.questionmark.com/l/26732/2014-07-25/8lhwf

Sindre, G., & Vegendla, A. (2015). *E-exams and exam process improvement*. Paper was presented at the UDIT/NIK 2015 Conference, Ålesund, Norway.

Sorensen, E. (2013). Implementation and student perceptions of e-assessment in a chemical engineering module. *European Journal of Engineering Education*, *38*(2), 172–185. doi:10.1080/03043797.2012.760533

Sullivan, G. M., & Artino, A. R., Jr. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, *5*(4), 541–542.

Tisdell, C. C. (2017). Critical perspectives of pedagogical approaches to reversing the order of integration in double integrals. *International Journal of Mathematical Education in Science and Technology*, *48*(8), 1285–1292. doi:10.1080/0020739X.2017.1329559

Tisdell, C. C. (2018a). Pedagogical alternatives for triple integrals: Moving towards more inclusive and personalized learning. *International Journal of Mathematical Education in Science and Technology*, *49*(5), 792–801. doi:10.1080/0020739X.2017.1408150

Tisdell, C. C. (2018b). On Picard's iteration method to solve differential equations and a pedagogical space for otherness. *International Journal of Mathematical Education in Science and Technology*. Advance online publication. doi:10.1080/0020739X.2018.1507051

Trochim, W. M. K. (2006). *The research methods knowledge base*. Retrieved from https://www.socialresearchmethods.net/kb/positvsm.php

von Grünigen, D., Benites de Azevedo e Souza, F., Pradarelli, B., Magid, A., & Cieliebak, M. (2018). Best practices in e-assessments with a special focus on cheating prevention. *Proceedings of 2018 IEEE Global Engineering Education Conference* (pp. 893–899). Washington, DC: IEEE.

Wang, V. C. (2009). *Handbook of research on e-learning applications for career and technical education: Technologies for vocational training*. Hershey, PA: IGI Global.

Wibowo, S., Grandhi, S., Chugh, R., & Sawir, E. (2016). A pilot study of an electronic exam system at an Australian university. *Journal of Educational Technology Systems, 45*(1), 5–33. doi:10.1177/0047239516646746

Author Biographies

Ananthan Ambikairajah is a passionate neuroscientist and educator, who has completed his Bachelor of Science (Neuroscience) and Master of Teaching (Secondary) degrees at the University of New South Wales, Sydney. Ananthan currently works with the Australian National University and is completing his PhD, which aims to understand the interaction between cardiovascular risk factors and menopausal status on structural brain changes. Ananthan has a keen interest in conducting research that can have a positive, widespread impact on society, in addition to communicating his scientific work to the public to spark interest, critical thought, and encourage scientific discourse.

Chris Tisdell is a professor of Mathematics at University of New South Wales and an honorary professor of Educational Research at The University of Queensland. His contributions to research aim to provide humanity with alternative constructs to work with that illuminate our world in a new way. His research portfolio is of a multidimensional nature, resting on four complementary pillars: (a) mathematics research, (b) STEM education research (particularly mathematics education research), (c) digital education, and (d) research into higher education.