Towards authentic e-Exams at scale: robust networked Moodle

Mathew Hillier
Monash University, Australia

Scott Grant
Monash University, Australia

Martin Coleman
Monash University, Australia

In this paper we present the design and user evaluation of a resilient online e-Exam platform that is capable of working without a network for most of the exam session, including the conclusion of an exam, without loss of data. We draw upon the education and technology acceptance literature as a basis for evaluation. The technology approach takes advantage of the Moodle learning management system quiz module as a means to provide an electronic workflow for assessments and builds on a range of open source components to construct the robust solution. The approach also enables rich, constructed assessment tasks by providing authentic 'e-tools of the trade' software applications and a consistent operating system on each student's BYO laptop. The robust Moodle exam deployment was trialled in two undergraduate units (subjects) at an Australian university. Students undertook a sequence of practice, mid-term and a final examination using the platform. Additional software and audio files were utilised as part of the exams. Student feedback on their experience was collected using pre and post surveys covering a range of issues related to technology acceptance.

Keywords: Keywords. e-Exams, networking, resilience, offline, authentic assessment.

Background

This paper focuses on the design and evaluation of a 'robust' online edition of the e-Exam platform developed in the third phase of work being carried out under an Australian Government funded project (TEAA, 2015; Fluck & Hillier, 2016). The project is looking at ways to modernise supervised examinations within the Australian higher education context where the primary aim has been to enable authentic assessment. We use the term 'e-Exam' (eExam) to specifically refer to a "timed, supervised, summative assessment conducted using each candidate's own computer running a standardised operating system" (Fluck & Hillier, 2017). This definition is in contrast to many existing computerised testing systems (QuestionMark, Examsoft, TCExam etc) that use single applications or web pages that provide a limited 'form' based environment for questions and responses. Such approaches add little to the design of exam-based assessments because the rich affordances of complex software as mind tools (Jonassen, 1991) for problem solving is not available to task designers or students.

We have previously discussed a set of requirements for an approach to e-Exams (Hillier & Fluck 2013) that has included enabling authentic assessment using e-tools of the trade, integrity, reliability and scalability that would ethically use student's bring-your-own laptop.

A primary aim of our work has been to allow technology to be a tool that enables the redefinition (Puentedura 2003, 2006) of assessment tasks, targeting higher order thinking (Krathwhol, 2002), rather then just replicate paper-based question formats in a digital form. We are working towards providing a holistic digital 'authentic assessment' (Crisp, 2009). By providing sophisticated software applications we can open up the pedagogical landscape of the exam room enabling assessment designers to set complex tasks that are better reflective of the employability needs of 21st century graduates, requiring a higher degree application, analysis, synthesis and evaluation.

Work carried out under the first two phases of the project focused on providing a transition between paper and digital exam scripts. In phase one, students were given a choice of handwriting or typing their responses using a word processor. Phase two saw the use of a range of 'e-tools of the trade' where constructed responses were enabled through the provision of multimedia, software development, diagramming and language translation tools. Responses were returned as a word processor document, program script or similar digital artefact.

The first two phases of our e-exam work did not utilise a network during the exam session itself. This minimised the requirements for complex infrastructure in the exam room and minimised risks associated with relying on networked servers for the duration of the e-exam event. However, this came at the cost of administrative

efficiency because data had to be loaded and retrieved from USB sticks for each exam.

The key stakeholder group that will be most impacted by the adoption of computers for exams is the students. As the primary users of the system we asked their opinions regarding ease of use, reliability, security and suitability of the e-Exam system via surveys and focus groups. Responses previous trials have been positive with students rating their experience above 4 on a 5-point scale across most measures. Full results of the earlier phases are reported elsewhere (Hillier, 2015b; Hillier & Lyon, 2018a,b).

What has become apparent from our trials of e-exams is that students were far more conservative in their views when faced with the idea of e-exams. Their level of enthusiasm was found to vary by discipline area (Hillier, 2014). This was unexpected given the high rate of use and ownership of computers amongst university students (McManus 2012). When given the opt-in choice of typing or handwriting the actual rates of adoption by students varied greatly with low figures ranging from 5% to 36% in some early e-exam trials (Hillier 2015b). In the UK, Purcell, Paterson & Mogey (2012) noted a 10% uptake in their e-Exam trials. Refinement of induction strategies saw this rise to around 50% in more recent opt-in Australian trials (Hillier & Lyon, 2018a). The most prevalent concern reported by students was the reliability of using computer technology in a time constrained, stressful, high stakes exam environment (Hillier 2014). This means that reliability needs to be demonstrably high to lessen unnecessary stress. The ideas of ‘graceful degradation’ and ‘progressive enhancement’ (W3C 2015) of capabilities and logistical continuity strategies (procedures) led us to implement multi-layered technical fall-backs. This is important so that if things go wrong it doesn’t result in a catastrophic failure. Using technology such as a networked learning management system (LMS), online web application (e.g. TCExam, QuestionMark Perception), or online remote desktop style environments (e.g. Citrix virtual desktop) typically have acceptable reliability levels when used for non-time critical activities (i.e. formative learning, out-of-class study, self-check quizzes or long duration project work). A key feature of these systems is the use of live networks for the entire assessment workflow or session. This is great for administrative efficiently and the provision of a unified toolset, but reliance on an active network connection to a server for the duration of the exam is also their "Achilles heel" that represents a single point of failure for the whole cohort of students. Recent examples of failures in networked computerised exams attest to this risk (Donovan, 2014; Peregoodoff, 2015; Strauss, 2016; Aubusson & Noyes 2018a, b). During the time-limited exam session any failures that impact the whole cohort are magnified. This can include the failure to start an exam, significant interruptions to working time during an exam, or aborted exam sessions. This leads to increased stress for already stressed students and disruptions to tight exam schedules.

These tales of trouble prompted us to consider how we can leverage the utility of a networked assessment system whilst minimising the risk of technical failure.

Research Questions

We have seen success in the first two phases of the project on the primary goal of our project - enabling a greater degree of authentic assessment within supervised exam spaces. However, with Universities in Australasia each regularly running 100,000 to 300,000 exam sittings annually (Roach 2017), to be ‘doable’ in this context, the technical and procedural approaches also need to be efficient. Therefore, a key guiding research question for the third phase of our work was:

"How can we maintain and extend a capability for higher order, rich, authentic, e-tools of the trade assessment tasks, while taking advantage of the administrative efficiencies of a network server but ensuring that the system was robust during the critical period of the exam event?"

Within a context where Learning Management Systems (LMS) still dominate learning environment and while striving for more authentic assessment, two specific questions were posed:

1) "How can we leverage the administrative efficiencies of a networked LMS for use in e-exams?"

And, at the same time:

2) "How can we minimise the risks associated with the reliance on a live network during the exam?".

Finally, we need to ensure that the primary users of the system, that is the students, accept the system for use under exam conditions:

3) "Do students accept the networked edition of the e-exam system as fit for purpose for undertaking supervised time limited exams?

Technical development approach

Our technical development approach has been component-based software development (CBSD). This is increasingly being used to deliver solutions that involve the re-use and adaption of what equate to 'virtual Lego bricks'. This includes both open-source software (Capiluppi, Boldyreff & Stol, 2011) and commercial off-the-shelf (COTS) products. The third phase of our work on the 'robust' networked approach to e-exams drew upon a range of open-source software such as Linux, Moodle and Safe Exam Browser (SEB) that we built upon, combined and configured. A small percentage of the overall code was custom developed but this was critical in ensuring the solution met our design goals. COTS hardware components such as laptops, servers, networking, USB storage devices and large USB hubs were also used. The rationale for using open source development for e-exams has previously been articulated in (Hillier & Fluck, 2013) while the use of COTS hardware was a practical and financial decision.

While there are many approaches and software systems used for computerised exams (Hörnblad & Brenner, 2015) none have combined the separate parts and concepts to create the set of features and capabilities that we present in this paper. Here we focus on the prior work that had a direct influence on our solution. The first of these is the open source Linux operating system. Using a 'Live' Linux approach to boot laptops means that laptops from Apple and those that normally run 'Windows' will instead run an identical software tool set. Its open source status provides a testing authority with the means to provide a known, consistent software environment to all exam candidates regardless of the native operating system on the laptop. A Live Linux system has also been used as an exam environment by others (Frankl, Schartner & Zebedin, 2012; Alfreosson, 2014; Britschgi, 2015), some of which we have previously reviewed (Fluck et al 2017). In a number of other cases Live Linux has been used in conjunction with BYO laptops to provide a controlled software environment, either by starting the laptop from a CD-ROM (Fluck, Pullen & Harper, 2009), then USB stick (Alfreosson, 2014; Lattu, 2014; Britschgi, 2015; Hillier & Fluck, 2017; Yioppilastutkinto, 2018) or a from network source (Frankl, Schartner & Zebedin, 2012). Several feature the use of a lock-down browser to serve as a gateway to a networked quiz or LMS server (Adesemowo, Johannes, Goldstone & Terblanche, 2016). The open source Safe Exam Browser (SEB) has proven successful in deploying computer lab based e-exams (Al Nadabi, 2015) and when used with BYO laptops, although instances of the latter are less common. SEB also has the ability to allow third party applications during an exam when specified, although this relies on separate provision of such applications by the owner of the equipment. In at least one case SEB, Moodle and Linux have been combined however the proponents of the approach admit that they are still reliant on a live connection for the duration of the exam (Frankl, Schartner & Jost, 2017). Few existing solutions allow for 'offline' use. One that does is the commercial closed source Examsoft product. However, Examsoft does not allow regular software applications to be used alongside quiz centric tools and therefore this limits the 'authenticity' of the assessment.

Moodle is a popular LMS used around the world and is open source allowing for easy inclusion. Moodle already has a very comprehensive question engine, a variety of question types and the ability to assess short text responses by pattern matching. Moodle can also allow the submission of files within the quiz itself which means that a wide range of software applications can be deployed as problem solving tools within the same system. This provides the opportunity for a mixed mode exam that includes selected response, convergent response and complex constructed responses. We also drew upon the large library of contributed code components including by Ward and Pinna (2018) for streamlined course enrolment, Hunt (2015) for offline capabilities and Hunt (2018) for Safe Exam Browser key pairing integration. We subsequently made modifications to these components to make them work more transparently with Safe Exam Browser and within the Live Linux environment. All of these factors meant that Moodle was an attractive for use as part of an e-Exam environment from the point of view of enabling authentic assessment within a robust e-workflow.

e-Exam platform capabilities

The e-Exam platform uses a combination of techniques and technologies. Detailed technical features of the customised Live Linux based e-Exam platform are explained further elsewhere (Hillier & Fluck, 2017). But briefly, we have deployed a customised Live Linux distribution loaded on multi-partition USB flash storage devices to boot a variety of student bring-your-own laptops. Customisations to the Linux OS includes restrictions on what the user can do and access within the system. Host drive access and third-party device access is prohibited as is access to a number of communications channels. Networking can be enabled and used in a controlled manner. A range of authentic software applications can be provided within the exam environment, including a full office suite, multimedia players, drawing tools and discipline specific tools, for example, Mathematics software (Maxima, R, Scilab, GeoGebra, Gummi LaTeX editor and NetLogo). In this phase of our work we have extended the ability of the platform to allow these tools to be used alongside a LMS

quiz. This creates a comprehensive mixed mode e-assessment platform. A variety of configurations are possible:

a) **Fully offline**: No network was used during the exam and all material is pre-loaded on the USB stick prior to the exam session. A suite of applications is made available on the USB with all work conducted and saved to USB storage. An office suite is used as the response composition environment and has a fully automatic document save feature. Over 20 trials have been conducted using this mode of operation. In phase one, paper-equivalent, student choice exams were run. Phase two saw post-paper trials where all students used a computer to complete tasks (e.g. computer programming, spread sheet, multimedia).
b) **Fully online**: The USB acts as a thin client secure gateway to a learning management system server or virtual desktop environment. A full suite of applications can be optionally provided via the USB as local working space. A highly reliable network and server infrastructure is required. The trade-off between administrative convenience and in-room reliability is that a network outage will result in a halt to the exam.
c) **Cached online**: The e-Exam USB provides a secure client that connects to a LMS, in this case a Moodle server using a key pair with SEB (the key prevents unauthorised access). Exam content is cached (from a Moodle quiz) at the start of the exam. The network is then optional and serves as an administrative convenience from then onwards. Auto synchronisation of student responses to the server and automatic fall-back to encrypted local storage make this approach more robust than standard online exams. A full suite of applications can be provided via the USB as local working space. Any files produced by a student can be submitted via Moodle or saved to the USB stick.

Detail of the process used to run a cached online e-exam is represented in Figure 1 and described below. It is worth noting that both the server and the client are fault tolerant and flexible. The e-Exam client USB hold a cache of the exam content and auto save or back-up the student's responses according to network conditions. The server can be run on common web infrastructure or from a Live USB stick (e.g. booted from a laptop in the exam room). In the latter case additional internal backup and automatic recovery for the server have been implemented in case of a server crash. Should this occur during an exam the clients continue to operate in off-line mode. When the server resumes operation, clients will automatically reconnect to the server and resume synchronisation of response data without loss of the user session.

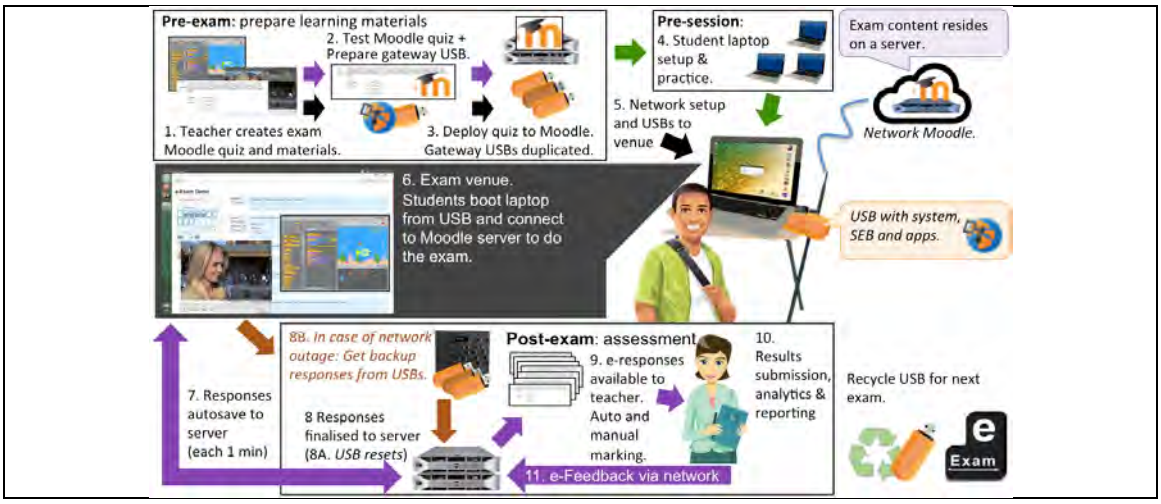

Figure 1: workflow for cached online e-exams with robust Moodle

The process for conducting an online e-Exam using the 'cached online' mode is outlined below:

1. Exam materials are prepared as a Moodle quiz including any file attachments.
2. Each set of exam materials is checked. The e-Exam USB sticks are configured and can be generic or customised for an exam or used to pre-load large resource files such as a long video. Under most common use the USBs would be configured with a generic set of applications and will be reusable.
3. Each set of exam materials is deployed to Moodle server. The e-Exam USB sticks are duplicated (or reused from previous sessions).
4. Students attend a practice session to check laptop compatibility and familiarise with the system.
5. Exam rooms are set up a 'quick start' instruction card and e-Exam USB sticks placed on desks instead of paper exam scripts. Power sockets were provided for each student. Spare laptops and USB WiFi dongles are on hand should they be required.
6. The exam session runs:
   a. After students enter the exam room they boot their laptop using the USB stick.
   b. Upon reaching the e-Exam system desktop a unique background image is displayed (serving as a visual

check for exam invigilators) along with the first user prompt.
   c. The system requests the user to provide their information: student id and name.
   d. The system then requests their network login (or automatically connects) to the network.
   e. Upon connection, the e-Exam system launches SEB to the preconfigured Moodle server.
   f. The student logs into Moodle.
   g. A link to the exam for this student (Moodle quiz) will be shown.
   h. If the SEB client and Moodle server keys match then the 'Attempt quiz now' button is displayed.
   i. Upon starting the quiz, the question content is cached to the client including images and multi-page quiz.
   j. A full suite of applications can be made available on the USB stick with complex constructed responses possible using the USB as local working space. Such responses can then be submitted via the Moodle file upload question type or saved to the 'answers' partition of the e-Exam USB.
   k. At the conclusion of the exam. The 'submit all and finish' button is used as per standard Moodle quiz submission. The student's responses are finalised and stored on the Moodle server.
   l. If a network connection is absent then an onscreen notice is shown and the attempt data submission action is redirected to local storage as an encrypted file (on the e-Exam USB).
   m. Student closes the software and shuts down the computer.
   n. The e-Exam client USBs are collected, counted and then students can leave the exam room.
7. Note: During the exam the client autosaves in-progress attempt data to the server every minute. However,
   a. If the network is absent, then each attempt data autosave action is redirected to local storage as an encrypted file (to e-Exam USB). The cached quiz allows the student to continue working.
   b. Upon network reconnection the client will re-sync attempt data to the Moodle server.
8. Responses are finalized to the server.
   a. If the student's responses were successfully submitted to the server then USB sticks are cleaned.
   b. If the student's responses could not be submitted online to the server then staff must undertake a data retrieval process via a large hub. The e-Exam admin tool is used to bulk retrieve attempt data files from USBs. The attempt files are then uploaded to the Moodle server.
9. The normal e-workflow for assessment within Moodle resumes with responses made available to the teacher for automatic and/or manual grading.
10. Standard Moodle features can be used to manage evaluation, analytics, reporting and feedback.

The focus in the remainder of this paper is on the evaluation of the 'cached online' mode of operation.

Study context

A round of trials were conducted in two undergraduate Chinese studies units, one at first year and one at third year level, at an Australian university with a sequence of three exams in each unit. The students participated on a voluntary basis either typing or handwriting with the majority of typists trying an e-exam for the first time. We were primarily seeking findings related to system usability and acceptance where both unit groups (first year and third year) used the same technology with tests in similar circumstances and the same discipline area. A key difference in the exam conditions in the two units was two offline dictionary applications were made available for use in the e-Exam system for the third-year unit, but not in the first-year unit. In the first-year unit short listening items were included in exams two and three that utilised audio headsets. All exams were administered locally in-class rather than in centrally run exam halls. Power was provided at each desk and WiFi was used instead of wired connections. A separate Moodle server that included plugins and our modifications was used instead of the institutional Moodle server.

Research methods used for evaluation

In evaluating the robust networked e-Exam solution we draw upon information systems theories around technology acceptance (Davis, 1989; Farzin, 2017) and the work of prior researchers (Dermo, 2009) who have evaluated e-assessment systems. Drawing in our previously articulated requirements and work by Dermo provided the basis for the evaluation survey instruments (See table 1 for questions).

Evaluation procedure

Early in the semester a call for expression of interest was announced to students in each unit and informed consent collected according to the approved ethics protocol. The voluntary nature of the trial meant that students could withdraw or change their mind at any stage during the process.

A practice session was held two weeks prior to first exam. The aim was to check that student's laptops were compatible with the system and to provide a preview of the exam format, the boot processes and software

environment. Observation notes were taken and a pre-exam survey was carried out that collected information on hardware that was used, issues encountered and student's perceptions of the system and processes. A preliminary analysis was done to capture any concerns and to address any technical issues.

Mid semester exams (exams one and two) were held using a Moodle quiz of constructed and selected response questions. Post surveys were completed to capture the student's experience of each session.

At the end of the semester a final exam (three) was held using a similar range of question types and software tools. The post survey was again used to collect student's perceptions of the experience.

In our analysis in this paper we focus on questions related to reliability and usability of the e-Exam system on the part of those that typed the exam.

Data analysis

Statistical analysis of Likert item survey questions (strongly agree 5, neutral 3 and strongly disagree 1) as shown in Tables 1 to 4 was done using SPSS v24 using alpha level of .05. Likert data pertaining to student's opinions were analysed item by item (not as a scale) and as such were treated as non-parametric (Jamieson, 2004). The statistical techniques we used included Mann & Whitney's (1947) U test to check the variance of two groups by study unit and by gender. The study by Dermo (2009) supports the choice non-parametric tests such as Mann & Whitney's test in analysing students' perceptions of e-assessment systems. We looked at pre-post paired Likert items using the Wilcoxon Signed Ranks Test (Wilcoxon, 1965). In cases where items did not meet the assumption of a normal distribution of differences when checked with a Shapiro & Wilk (1965) test ($p = .05$) we used the Sign Test (Roberson, Shema, Mundfrom, & Holmes, 1995). We selected the second exam as post comparison point because this was considered the most settled set of events. The first exam represented first real use and final exams in each unit encountered some organisational difficulties that were likely to influence results. We also acknowledge that the relatively small sample sizes could have impacted the accuracy of the results due to the dropping of 'ties' in the analysis (Mendenhall, Wackerly & Scheaffer, 1989). We also used the Friedman test (1939) to examine if differences existed over a time series of measures (i.e. a pre and a sequence of three post-tests). Selected survey items that had shown significant results in the previous analysis step for the time series test were analysed in pairs. Missing responses were excluded on a pair-wise (test-by-test) basis. We agreed with Dermo (2009) in considering that opinion data did not represent an objective truth about the e-exam system, but rather statistical results are indicative of the strength of the opinions of this particular group of students. Similarly, the study is limited in that students were not randomly assigned to typing or handwriting conditions and relatively small samples mean that the statistical findings cannot be generalised beyond this study.

Findings

In the first-year unit a soft target of 30 typing places were offered to the class of 124 with 20 opting to attend the practice session and 14 typing the final exam. In the third-year unit e-exams were notionally run on an opt-out basis, with 29 (94%) doing the practice session and 27 (87%) typing the final exam out of 31 enrolled.

Following the series of exams and practice sessions we compared the pre and post survey responses shown in Table 1. Overall, students reported a positive experience of using the Moodle based e-exam. Question 10 relating to usability rated well and impressions of reliability (Q7 and Q8) increased overall (see Table. 1). When we compared student's pre and post opinions using a 'Sign test'. Missing responses were excluded on a pair-wise (test-by-test) basis. Three items relating to perceptions of reliability saw a significant increase in agreement from pre to post ($p =< 0.05$).

We investigated differences in opinions between units because units were at different year levels in the degree program, allowed different software to be used and experienced different logistical conditions. A Mann-Whitney U test was used to explore the opinions of 19 individuals from the first-year unit and 28 individuals from the third-year unit. Some differences between grouped responses were found, particularly later in the sequence of exams (See Table 2). The second exam in each sequence was used for post.

Table 1: Aggregated Likert Item Results for Pre and Post (Exam two)

Likert items	**Pre**			**Exam 2**			**M**	**Sig.**
(Strongly Agree 5, Neutral 3, Strongly disagree 1)	*n*	*M*	*SD*	*n*	*M*	*SD*	Diff.	*p.*
1) My laptop is reliable for use in a computerised exam.	38	3.0	1.3	37	3.4	1.4	0.4	**0.05**[b]
2) I can use the e-Exam system just as well as my own laptop.	38	3.2	1.3	36	3.6	1.2	0.4	0.14[a]
3) It was easy to answer multiple-choice questions.[c]	16	4.3	0.6	25	4.4	0.6	0.1	0.38[b]
4) The included software was useful.	36	4.1	0.9	38	3.7	1.0	-0.3	0.42[a]
5) Moodle worked well as an exam environment.	36	4.1	0.6	39	4.2	0.6	0.1	0.55[b]
6) I am concerned about network outages impacting my exam.	36	3.3	1.2	39	3.1	1.3	-0.2	0.45[a]
7) I felt the e-Exam system was reliable against technical failures.	38	3.1	1.1	37	3.8	0.8	0.7	**0.01**[a]
8) I am reassured the e-Exam system was robust against network outages.	35	3.4	0.7	39	3.8	0.7	0.4	**0.03**[a]
9) I feel the e-Exam System is secure against cheating.	38	4.2	0.7	37	4.2	0.7	0.0	0.77[b]
10) Overall, I feel the e-Exam System is easy to use.	38	4.0	0.9	37	4.1	0.7	0.1	1.00[b]
11) I now feel relaxed about using the e-Exam system for exams.	38	3.6	1.0	37	4.0	0.8	0.3	0.50[b]
12) I would like to use a computer for exams in the future.	38	3.8	0.9	38	3.9	0.9	0.1	1.00[b]
13) I would recommend the e-Exam system to others.	38	3.6	0.9	37	4.0	0.7	0.4	0.33[b]

[a.] Wilcoxon Signed Ranks Test used. [b.] Sign Test used. [c.] This question was labelled 'skip of not applicable'.

Table 2: Likert Item Results by Unit for Pre and Post (Exam two)

Likert items	1st yr unit		3rd yr unit			
(Strongly Agree 5, Neutral 3, Strongly disagree 1)	*mdn*[c]	*M*[b]	*mdn*[c]	*M*[b]	*U*[a]	*p*
Pre 1) My laptop is reliable for use in a computerised exam.	3	3.2	3	2.9	146.5	0.51
Pre 2) I can use the e-Exam system just as well as my own laptop.	3	2.9	3	3.3	138	0.35
Pre 3) It was easy to answer multiple-choice questions.	4.5	4.5	4	4.2	22.5	0.36
Pre 4) The included software was useful.	4	3.5	4	4.3	77	**0.03**
Pre 5) Moodle worked well as an exam environment.	4	4.1	4	4.1	137	0.99
Pre 6) I am concerned about network outages impacting my exam.	4	3.5	4	3.3	123	0.61
Pre 7) I felt the e-Exam system was reliable against technical failures.	3	2.9	3	3.3	134	0.29
Pre 8) I am reassured the e-Exam system was robust against network outages.	4	3.4	3	3.4	125	0.79
Pre 9) I feel the e-Exam System is secure against cheating.	5	4.5	4	4.0	100.5	**0.02**
Pre 10) Overall, I feel the e-Exam System is easy to use.	4	3.7	4	4.2	143.5	0.41
Pre 11) I now feel relaxed about using the e-Exam system for exams.	4	3.5	4	3.8	161.5	0.84
Pre 12) I would like to use a computer for exams in the future.	4	4.2	4	3.6	98.5	**0.03**
Pre 13) I would recommend the e-Exam system to others.	3.5	3.6	4	3.7	160	0.80
Post 1) My laptop is reliable for use in a computerised exam.	5	3.6	4	3.3	125	0.31
Post 2) I can use the e-Exam system just as well as my own laptop system.	4	3.9	4	3.4	104	0.09
Post 3) It was easy to answer multiple-choice questions.	5	4.8	4	4.0	28.5	**<.01**
Post 4) The included software was useful.	3	3.5	4	3.9	125	0.23
Post 5) Moodle worked well as an exam environment.	4.5	4.5	4	4.0	105	**0.02**
Post 6) I am concerned about network outages impacting my exam.	2	2.5	4	3.5	101.5	**0.03**
Post 7) I felt the e-Exam system was reliable against technical failures.	4	4.2	4	3.6	94.5	**0.03**
Post 8) I am reassured the e-Exam system was robust against network outages.	4	4.1	4	3.6	102	**0.02**
Post 9) I feel the e-Exam System is secure against cheating.	4	4.2	4	4.2	153	0.75
Post 10) Overall, I feel the e-Exam System is easy to use.	5	4.5	4	3.9	87.5	**0.01**
Post 11) I now feel relaxed about using the e-Exam system for exams.	5	4.5	4	3.6	63	**<.01**
Post 12) I would like to use a computer for exams in the future.	5	4.7	3	3.5	37.5	**<.01**
Post 13) I would recommend the e-Exam system to others.	5	4.6	4	3.7	56	**<.01**

[a.] Mann Whitney U Test to compare the first and third year level units. [b.] M= mean. [c.] mdn = median.

Items of particular interest were perceptions of system reliability Q7 "I felt the e-Exam system was reliable against technical failures" and Q8 "I am reassured the e-Exam system was robust against network outages". Our initial analysis yielded significant changes between pre and post exam two (Table 1). We further investigated differences within units between each of the four events using a time series. Table 3 displays overall Friedman tests (F) for each series with only the first-year unit showing significant results. Post-hoc Wilcoxon tests (T) for adjacent pairs in the series show only the pre to post 1 adjacent pairs showed significant changes for Q7 (means and standard deviations are provided for clarity).

Table 3: Time series on perceptions of system reliability

Question	**Pre**		**<– T –>**		**Post 1**		**<– T –>**		**Post 2**		**<– T –>**		**Post 3**		**F**	
Q7 I felt the e-Exam system was reliable against technical failures	*M*	*SD*	*Z*	*p*	*M*	*SD*	*Z*	*p*	*M*	*SD*	*Z*	*P*	*M*	*SD*	χ2(2)	*p*
First year unit (n=8)	2.9	1.3	-2.226	**0.03**	4.1	0.8	1.00	0.32	4.2	0.8	0	1.00	4.1	0.9	11.526	**0.01**
Third year unit (n=17)	3.3	0.9	-2.111	**0.04**	3.7	0.8	0	1.0	3.6	0.7	-1.89	0.06	3.4	0.8	5.885	0.12
Q8 I am reassured the e-Exam system was robust against network outages	*M*	*SD*	*Z*	*p*	*M*	*SD*	*Z*	*p*	*M*	*SD*	*Z*	*P*	*M*	*SD*	χ2(2)	*p*
First year unit (n=6)	3.4	0.8	-1.633	0.10	4	0.9	-.577	0.56	4.1	0.7	-1	0.32	4.3	0.9	9.811	**0.02**
Third year unit (n=19)	3.4	0.6	-.277	0.78	3.5	0.8	-.351	0.73	3.6	0.6	-.277	0.78	3.7	0.6	2.333	0.51

We also examined if student's declared future use intentions and recommendations may have changed after each contact with the e-exam system. A time series was done by unit with respect to Q12 "I would like to use a computer for exams in the future" and Q13 "I would recommend the e-Exam system to others". Although some minor movements of ratings were observed over the time series these were not found to be statistically significant after running a set of Friedman Tests within each unit.

Finally, we reviewed the technical issues that arose during the trials. At each stage typists were asked "Did you experience any technical difficulties during this session?" A summary of the technical issues encountered during pre (practice) and post (exams) is shown in Figure 2.

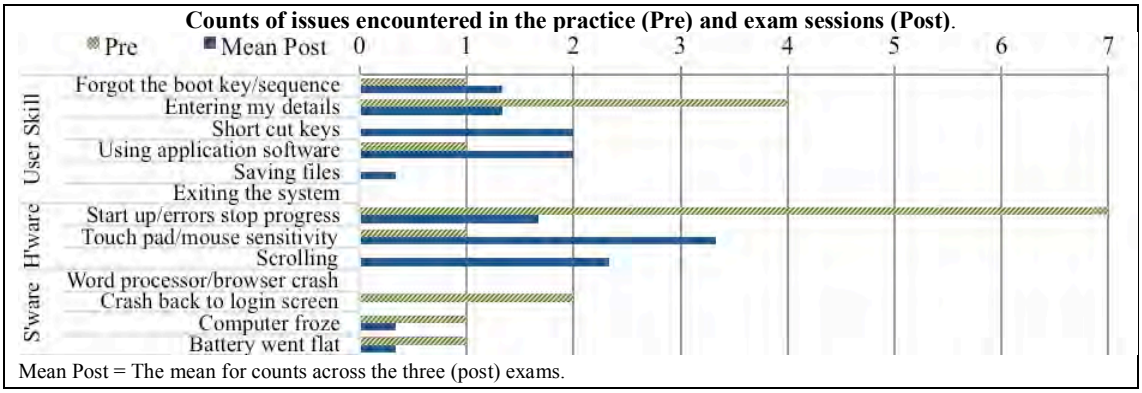

Figure 2: Reported technical issues

Discussion

Overall students were happy with the e-Exam system and agreed that they would recommend it to others (Table 1). Students agreed it was fit for purpose in terms of ease of use and reliability. However, it became apparent that lower ratings were given by the 3rd year unit group in post surveys. In many cases the ratings between the two-unit groups were significantly different (Table 2). Similarly, the ratings given for reliability items in the three post surveys (Table 3) were also lower from the third-year unit. The split in opinion could have been due to technical usability problems with the 3rd party 'Dim Sum' dictionary application used only in the third-year unit. We observed students frequently using it in solving exam questions and its problematic nature featured strongly in student verbal and written comments during exam sessions. It is therefore possible this coloured the 3rd year student's overall impression of the e-Exam approach. The final round of exams in both units suffered from some logistical failings on the part of the organisers and this may have also coloured participant's impression of the process. A dip or levelling off of mean reliability ratings can be seen in the third exam (Table 3).

Overall, the agreement scores and those from the first-year unit (which was not afflicted by the dictionary application) were in line with previously reported e-exam studies using a spreadsheet (Hillier & Grant, 2018) and word processor documents (Hillier & Lyon 2018a).

In regard to technical issues as shown in figure 2, those issues that related to user familiarity were linked to processes (i.e. use of non-standard user names on the test server, Apple users encountering the 'Windows' short cut keys used in the e-Exam system) and as previously mentioned, to poor usability of the 3rd party 'Dim Sum' dictionary application. Most of the hardware compatibility issues that could be considered 'blockers' were encountered in the practice sessions. If a student's own laptop was found to be incompatible then they were

offered the use of a loan laptop. Therefore, the practice sessions served the intended purpose of heading off significant problems before they got into the exam room itself. When the exams were run, all students who started an exam with the intention to type was able to do so. In some of cases this meant swapping the student's own laptop with an institution owned laptop therefore reinforcing the recommendation for contingency measures such as having spare laptops on hand. Minor (non-blocking) issues such as sensitive touchpads and scrolling were more prominently reported in the exam events. Wired mice that could have solved the issue were recommended but few students took up the offer. Perhaps not enough time was spent in the practice session to surface secondary issues that in turn became more problematic in the real exams. Software related issues were mainly overcome (through further software development and driver updates) by the time real exams were run. Two glitches occurred in one of the final exams that necessitated a computer restart (due to system freeze) and a quiz restart (the user found themselves outside the Moodle quiz for an unknown reason either due to a browser glitch or user error) but in both cases students were able to continue the exam without loss of response data thanks to the backup provisions in the e-Exam system. Finally, in another exam session, one laptop lost the WiFi connection to the server that became apparent at submission time (a notice was displayed on screen). Attempts at a manual re-connection were unsuccessful, however due to the e-Exam system redirecting data backup to USB storage all responses were saved and were later uploaded to the Moodle server. Students concerns over exam interruptions due to network outages (Q6) were alleviated as time went on with drops between pre and post ratings (evident in Table 1 and 2). Overall, all students who typed the exam were able to successfully complete with the protective measures built into the e-Exam system working to prevent data loss.

Conclusion

In this study we set out to see if we can leverage the administrative efficiencies of a networked LMS for use in e-exams whilst overcoming the risks associated with the reliance on a live network during the exam. From this point of view we have been successful. Both bench testing and a series of live trials have proven that the e-Exam platform running a 'robust' implementation of Moodle can survive both system crashes and network outages without loss of critical student response data.

We also found that students were able to accept the networked edition of the e-exam system as being fit for purpose for undertaking supervised time limited exams. Student's gave reliability and usability ratings in line with previous off-line e-exam trial outcomes.

Acknowledgments. The Australian Government Department of Education and Training for financial support. Special thanks to Tim Hunt, Open University UK and Daniel Schneider, ETH Zurich for their technical advice.

References

Adesemowo, A. K., Johannes, H., Goldstone, S. & Terblanche, K. (2016) The experience of introducing secure e-assessment in a South African university first-year foundational ICT networking course. Africa Education Review, 13(1), 67–86. https://doi.org/10.1080/18146627.2016.1186922

Al Nadabi, Z. S. H. (2015) Features of an online English language testing interface. In Globally connected, digitally enabled (pp. 369–373). Perth, Australia: Australasian Society for Computers in Learning in Tertiary Education. Retrieved from https://espace.library.uq.edu.au/view/UQ:386253

Alfreosson, F. (2014, September) Bring-your-own-device Exam System for Campuses. Presented at the 28th NORDUnet Conference, Uppsala University, Sweden. Retrieved from https://events.nordu.net/display/NORDU2014/Bring-your-own-device+Exam+System+for+Campuses

Aubusson, K. & Noyes, J. (2018, February 20) Burnt-out doctors deeply distressed by botched high-stakes exam. The Sydney Morning Herald. Retrieved from https://www.smh.com.au/national/nsw/burnt-out-doctors-deeply-distressed-by-botched-high-stakes-exam-20180220-p4z117.html

Aubusson, K. & Noyes, J. (2018, February 21) Company behind botched medical exam has track record of failure. The Sydney Morning Herald. Retrieved from https://www.smh.com.au/national/company-behind-botched-medical-exam-has-track-record-of-failure-20180221-p4z15a.html

Britschgi, V. (2015) The Finnish matriculation examination (presentation slides). https://www.ylioppilastutkinto.fi/images/sivuston_tiedostot/Kehittaminen/YTL_presentation_English_update.pdf

Capiluppi, A., Boldyreff, C. & Stol, K.-J. (2011) Successful Reuse of Software Components: A Report from the Open Source Perspective. In Open Source Systems: Grounding Research (pp. 159–176). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-24418-6_11

Crisp, G. (2009) Towards Authentic e-Assessment Tasks (Vol. 2009, pp. 1585–1590). Presented at the EdMedia: World Conference on Educational Media and Technology, Honolulu, HI, USA. http://www.editlib.org/p/31689/

Davis, F. D. (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Quarterly, 13 (3): 319–340, doi:10.2307/249008

Dermo, J. (2009) E-assessment and the student learning experience: A survey of student perceptions of e-assessment. British Journal of Educational Technology, 40, 203–214

DistroWatch (2018) DistroWatch, Unsigned Integer Limited, Copenhagen https://distrowatch.com/

Donovan, D. (2014, August 18) Wake Forest law grad sues ExamSoft over bar exam debacle – North Carolina Lawyers Weekly. North Carolina Lawyers Weekly. Retrieved from https://nclawyersweekly.com/2014/08/18/wake-forest-law-grad-sues-examsoft-over-bar-exam-debacle/

Farzin, S. (2017) Attitude of Students Towards E-Examination System: an Application of E-Learning. Science Journal of Education, 4(6), 222. DOI: 10.11648/j.sjedu.20160406.19

Fluck A. & Hillier, M. (2016) Innovative assessment with eExams. Australian Council for Computers in Education Conference, Brisbane, 29 Sep to 2 Oct. http://conference.acce.edu.au/index.php/acce/acce2016/paper/view/34/27

Fluck, A. & Hillier, M. (2017) eExams: Strength in Diversity. In A. Tatnall & M. Webb (Eds.), *Tomorrow's Learning: Involving Everyone. Learning with and about Technologies and Computing* (Vol. 515, pp. 409–417). Cham: Springer. https://doi.org/10.1007/978-3-319-74310-3_42

Fluck, A. (2007) Can tertiary eAssessment change secondary school cultures? In F. Khandia (Ed.), 11th CAA International Computer Assisted Assessment Conference (pp. 191–203). Loughborough, UK: Loughborough University. http://caaconference.co.uk/pastConferences/2007/proceedings/Fluck%20A%20f2_formatted.pdf

Fluck, A., Pálsson, H., Coleman, M., Hillier, M., Schneider, D., Frankl, G. & Uolia, K. (2017). eExam symposium: design decisions and implementation experience. IFIP World Conference on Computers in Education, Dublin.

Fluck, A., Pullen, D. & Harper, C. (2009) Case Study of a Computer Based Examination System. Australasian Journal of Educational Technology, 25(4), 509–523.

Frankl, G., Schartner, P. & Jost, D. (2017) The "Secure Exam Environment": E-Testing with Students' Own Devices. In A. Tatnall & M. Webb (Eds.), *Tomorrow's Learning: Involving Everyone. Learning with and about Technologies and Computing* (Vol. 515, pp. 179–188). Cham: Springer. https://doi.org/10.1007/978-3-319-74310-3_20

Frankl, G., Schartner, P. & Zebedin, G. (2012) Secure online exams using students' devices (pp. 1–7). IEEE Global Engineering Education Conference (EDUCON), Marrakech. https://doi.org/10.1109/EDUCON.2012.6201111

Friedman, M. (1939) "A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance". Journal of the American Statistical Association. 34 (205): 109. doi:10.2307/2279169

Hillier, M. (2014) "The Very Idea of e-Exams: Student (Pre)conceptions", 31st ASCILITE conference, 24-26 Nov, Dunedin. http://ascilite.org/conferences/dunedin2014/files/fullpapers/91-Hillier.pdf

Hillier, M. (2015a) "e-Exams with student owned devices: Student voices" Proceedings of the International Mobile Learning Festival: Mobile Learning, MOOCs and 21st Century learning, 22-23 May, Hong Kong SAR China http://transformingexams.com/files/Hillier_IMLF2015_full_paper_formatting_fixed.pdf

Hillier, M. (2015b) To type or handwrite: student's experience across six e-Exam trials. Proceedings of the Australasian Society for Computers in Learning in Tertiary Education conference, Perth, Australia: ASCILITE. Retrieved from http://transformingexams.com/files/Hillier_2015_ascilite_fp.pdf

Hillier, M. & Fluck, A. (2013) Arguing again for e-exams in high stakes examinations. In H. Carter, M. Gosper, & J. Hedberg (Eds.), Electric Dreams (pp. 385–396). Macquarie University. Retrieved from http://www.ascilite.org.au/conferences/sydney13/program/papers/Hillier.pdf

Hillier, M. & Fluck, A. (2017) Transforming Exams - How IT works for BYOD e-Exams. Australasian Society for Computers in Learning in Tertiary Education conference (pp. 100-105). Toowoomba, 4-6 December. http://2017conference.ascilite.org/wp-content/uploads/2017/11/Concise-HILLIER.pdf

Hillier, M. & Grant, S. (2018) Using spreadsheets as an e-Exam system, Presentation at HERDSA conference, Adelaide.

Hillier, M. & Lyon, N. (2018a) Student experiences with a bring your own laptop e-Exam system in pre-university college, *Open Conference on Computers in Education* (IFIP TC3), 25-28 June, Linz, Austria.

Hillier, M. & Lyon, N. (2018b) Writing e-Exams in pre-university college, *Open Conference on Computers in Education* (IFIP TC3), 25-28 June, Linz, Austria.

Hörnblad, P. & Brenner, M. (2015). Digital Tentamen - Inkubator Teknisk Information. Retrieved 22 February 2018, from https://portal.nordu.net/display/Inkubator/Digital+Tentamen

Hunt, T. (2015) 'offlinemode', Moodle plugins. https://moodle.org/plugins/

Hunt, T. (2018) 'safeexambrowser', Moodle plugins. https://moodle.org/plugins/

Jamieson, S. (2004). Likert scales: how to (ab)use them. Medical Education, 38(12), 1217–1218. https://doi.org/10.1111/j.1365-2929.2004.02012.x

Jonassen, D. H. (1991) What are cognitive tools? In P. A. M. Kommers, D. H. Jonassen, & J. T. Mayes (Eds.), *Cognitive tools for learning* (pp. 1–6). Berlin, Germany: Springer-Verlag. https://doi.org/10.1007/978-3-642-77222-1_1

Krathwohl, D. R. (2002) A Revision of Bloom's Taxonomy: An Overview. Theory Into Practice, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2
Lattu, M. (2014) Finland's national matriculation exams goes electronic, Transforming Assessment, 12 September, http://www.transformingassessment.com/eAS_2014/events_12_september_2014.php
Mann, H. B. & Whitney, D. R. (1947) On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics, 18(1), 50–60. https://doi.org/10.1214/aoms/1177730491
McManus, M. (2012) New media Literacies and Personalized Learning, Key Note Address, Sixth U21 Educational Innovation Conference, 8-9 Nov, Singapore http://www.universitas21.com/RelatedFile/Download/387
Mendenhall, W., Wackerly, D.D. & Scheaffer, R.L. (1989) 15: Nonparametric statistics, *Mathematical statistics with applications* (Fourth ed.), PWS-Kent, pp. 674–679, ISBN 0-534-92026-8
Purcell, M., Paterson, J., & Mogey, N. (2012) Exams: Comparing Handwritten Essays with those Composed on Keyboards (Final Report). Heslington, UK: The Higher Education Academy. http://www.docs.hss.ed.ac.uk/divinity/About%20the%20school/Elearning/HEA2012-Final%20Report.pdf
Peregoodoff, R. (2015) Large-Scale BYOD Wireless LMS Midterm and Final Exams. Professional Development Seminar presented at the EDUCAUSE Annual Conference, 28 October. https://events.educause.edu/annual-conference/2015/proceedings/large-scale-byod-wireless-lms-midterm-and-final-exams
Puentedura, R. R. (2003) A Matrix Model for Designing and Assessing Network-Enhanced Courses, Hippasus. Retrieved 26 February, 2018, http://hippasus.com/resources/matrixmodel/index.html
Puentedura, R. R. (2006) Transformation, Technology, and Education, Hippasus. Retrieved February 26, 2018, from http://hippasus.com/resources/tte/
Roach, A. (2017) Exams Network 2017 Benchmarking Survey, Unpublished report.
Roberson, P. K., Shema, S. J., Mundfrom, D. J., & Holmes, T. M. (1995) Analysis of paired Likert data: how to evaluate change and preference questions. Family Medicine, 27(10), 671–675.
Shapiro, S. S. & Wilk, M. B. (1965) An analysis of variance test for normality (complete samples), *Biometrika* 52 (3-4): 591–611. doi:10.1093/biomet/52.3-4.591
Strauss, V. (2016, April 21). Pearson's history of testing problems — a list. Washington Post. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2016/04/21/pearsons-history-of-testing-problems-a-list/
TEAA (2015) Transforming exams across Australia, Australian Government Office for Learning and Teaching, Grant, ID15-4747 http://transformingexams.com
W3C (2015) Graceful degradation versus progressive enhancement, 2 January https://www.w3.org/wiki/Graceful_degradation_versus_progressive_enhancement
Ward, M & Pinna, R. (2018) 'autoenrol', Moodle plugins. https://moodle.org/plugins/
Wilcoxon, F. (1945) Individual comparisons by ranking methods. Biometrics Bulletin, 1(6): 80–83. doi:10.2307/3001968.
Yioppilastutkinto (2018) Digital Matriculation Examination, The Matriculation Examination Board Finland https://www.ylioppilastutkinto.fi/en/matriculation-examination/digital-matriculation-examination

Please cite as: Hillier, M., Grant, S. & Coleman, M. (2018). Towards authentic e-Exams at scale: robust networked Moodle. In M. Campbell, J. Willems, C. Adachi, D. Blake, I. Doherty, S. Krishnan, S. Macfarlane, L. Ngo, M. O'Donnell, S. Palmer, L. Riddell, I. Story, H. Suri & J. Tai (Eds.), Open Oceans: Learning without borders. Proceedings ASCILITE 2018 Geelong (pp. 131-141). https://doi.org/10.14742/apubs.2018.1919