2 March 2016: 07:00AM GMT

Designing for Learner Engagement with Computer-based Summative Exams
Richard Walker (E-Learning Development) and Zoe Handley (Department of Education), University of York, UK

Your Webinar Hosts
Professor Geoff Crisp,
PVC Education, University of New South Wales
g.crisp[at]unsw.edu.au

Dr Mathew Hillier,
Office of the Vice-Provost Learning & Teaching,
Monash University
mathew.hillier[at]monash.edu

Just to let you know:
By participating in the webinar you acknowledge and agree that:
The session may be recorded, including voice and text chat communications (a recording indicator is shown inside the webinar room when this is the case).
We may release recordings freely to the public which become part of the public record.
We may use session recordings for quality improvement, or as part of further research and publications.
Designing for learner engagement with computer-based summative exams

Richard Walker
E-Learning Development
Zoe Handley
Department of Education
University of York, UK
Graduate School, Department of Education, University of York

MA Applied Linguistic in (English) Language Teaching

MA Education

MA Global and International Citizenship education

MSc Psychology in Education

MA Social Justice and Education

MA TESOL
Drivers for developing computer-based testing

- Address the marking effort for a summative ‘pen-and-paper’ assessment for research methods (MA Language Learning & Teaching) for 150+ international students, involving MCQ & open questions - challenging during busy spring term teaching
- Improve the efficiency of exam delivery
- Improve the consistency of marking for MCQ questions
- Reduce the marking load – ensure sustainable teaching and marking workload
- Automate reporting: item analysis; difficulty & discrimination analysis (perform longitudinally)
- Inform our institution – scoping towards the development of a centrally supported computer-based assessment service
Our challenge

- Designing from scratch a computer-based assessment for research methods
  *(MA Language Learning & Teaching)*
- Developing an assessment to address criticality & higher order thinking
- Delivering an assessment successfully across multiple test venues simultaneously for 150+ PG international students, with no prior exposure to computer-based testing at the University
- Managing the assessment within a ‘greenfield’ institutional context, with no established policy / protocols in place
Our journey: An evolutionary approach

- QMP trials (2009-2012)
- VLE hosted testing & exam (2013-2014)
- Benchmarking & technology review (2012-2013)
- VLE exam as a supported service (2014-15)

Formative tests – low stakes

Formative & summative exams – high stakes

Technical review & dummy tests

Formative testing informing higher stakes exam

VLE hosted testing & exam (2013-2014)
Agenda for the webinar

- **Review of sector developments and studies on computer-based testing**
  - Focus on student attitudes towards high stakes computer-based testing

- **Description of our exam environment and assessment design**

- **Review of our research approach and key findings**

- **Transferability of our findings to different computer-based testing contexts**

- **Applying the lessons learned:**
  - Impact of our findings on current support procedures for high stakes computer-based testing
1. What do we know about computer-based testing?

Effective Assessment in a Digital Age
http://www.jisc.ac.uk/whatwedo/programmes/elearning/assessment/digiassess.aspx

Report on Summative E-Assessment Quality (REAQ):
http://www.jisc.ac.uk/media/documents/projects/reaqfinalreport.pdf
### High stakes testing across UK HE

<table>
<thead>
<tr>
<th>Year</th>
<th>100%</th>
<th>75%-99%</th>
<th>50%-74%</th>
<th>25%-49%</th>
<th>1%-24%</th>
<th>0%</th>
<th>Don’t know</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>2%</td>
<td>5%</td>
<td>4%</td>
<td>13%</td>
<td>62%</td>
<td>4%</td>
<td>10%</td>
<td>-</td>
</tr>
<tr>
<td>2012</td>
<td>0%</td>
<td>1%</td>
<td>4%</td>
<td>10%</td>
<td>62%</td>
<td>5%</td>
<td>9%</td>
<td>8%</td>
</tr>
<tr>
<td>2010</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>14%</td>
<td>60%</td>
<td>12%</td>
<td>12%</td>
<td>-</td>
</tr>
<tr>
<td>2008</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>4%</td>
<td>64%</td>
<td>16%</td>
<td>15%</td>
<td>-</td>
</tr>
</tbody>
</table>

2014 TEL Survey: n=96 responding institutions

Source: UCISA TEL Survey 2014

[http://www.ucisa.ac.uk/bestpractice/surveys/tel.aspx](http://www.ucisa.ac.uk/bestpractice/surveys/tel.aspx)
Most research has focused on the differential impact of computer-based vs. pen-and-paper assessments on student achievement (Kingston, 2008; Leeson, 2006; Mead & Drasgow, 2008)
– differences small and not of practical significance

Attention has been directed to the relationship between individual differences (gender, race, digital literacy) and performance on computer-based assessments
– effects of learner variables tend to be small (Leeson, 2006)

Typically this has been case study research / surveys with few post-examination surveys and focus groups
The postgraduate experience?

- Only a limited number of studies on students’ attitudes (e.g. Dermo, 2009; Hillier, 2014; Walker, Topping & Rodrigues, 2008)

- Studies have flagged computer anxiety; perceptions of computer self-efficacy; concerns of test security (Deutsch et al., 2012; Hillier, 2014; Ozden et al. 2004)

- Focus has been on the undergraduate experience, reporting on a specific testing intervention
  - missing a longitudinal picture of student attitudes to high stakes computer-based testing

- The literature is silent on the postgraduate taught experience
  - what do we know about this stakeholder group?
Discussion Point

- Do international postgraduate taught (PGT) students require specific support strategies to engage with high stakes computer-based testing?

- Do they raise distinct challenges (in comparison with other cohorts – e.g. undergraduates), influencing their likely engagement with computer-based testing?
2. Our VLE Exam environment

- Dedicated instance of institutional VLE (Blackboard Learn v. 9.1), separate from the main VLE platform
  - dedicated servers (scalable to support 200+ students);
    all traffic is exam traffic
  - system restart prior to exam to counter memory leaks; exam caches primed by running exam on each app server

- Secure: access controlled through IP restrictions & module availability
  - access for duration of exam only
  - separate user accounts (exam number / student ID)

- Locked-down through desktop environment using virtual client
  - boots straight into browser with access to VLE Exam server URL only
  - can be used anywhere on campus (multiple venues)
Our Teaching & Assessment Context

- 150 + international postgraduate taught students
- Modules in Research Methods, Planning and Communication Research, and Independent Study (Empirical Dissertation)
- Research Methods (Exam: 2 hours)
  - Assesses higher-order thinking skills, including criticality
  - Combined:
    - Multiple choice questions (75%) – automated marking
    - Open response / short essay questions (25%, i.e. 2 Qs) – hand marked
- First centrally-facilitated high stakes e-assessment exam at the University (2013-14)
Multiple-choice

The main advantage of using a self-created observation schedule rather than modifying an existing observation schedule is:

- The validity increases as it can be piloted and tested with participants you have pre-selected.
- It relates precisely to the particular research questions being asked in the project.
- It is likely to result in a higher grade being given by examiners as it shows originality.
- All of the above

Options:
- Show answers in random order, allow partial credit
Identify the dependent variables in the following research question.

Schauer (2006): Do learners in English as a foreign language and English as a second language contexts display differences in their recognition and

Select the dependent variables from the list below.

NOTE: There may be more than one correct answer.

- Recognition of pragmatic errors
- Ratings of pragmatic errors
- Ratings of grammatical errors
- English as a foreign language context
- Recognition of grammatical errors
- English as a second language context

Options:
- Show answers in random order, allow partial credit
Question 3

Look at the list below. Please indicate whether these would be found predominantly in quantitative or qualitative research.

NOTE: You may use each option from the drop-down list more than once.

- Interviews
- Questionnaires
- Focus groups
- Large sample size
- Observations
- Statistics
- Small sample size

Moving to another question will save this response.

- Options:
  - Allow partial credit
Fill in multiple blanks

Question 10

Please fill in the following blanks with appropriate terms.

NOTE: Please check your spelling.

In [ ] studies, data is gathered over an extended period of time, whereas in [ ] studies, one produces a [ ]

Options:

- Allow partial credit
- Contains | Exact match | Pattern match
- Case sensitive
Question 6

What do a questionnaire and a structured interview have in common? How do they differ?

Provide 2 similarities and 2 differences in the space provided.

Make sure your explain each point fully.
Revising the examination

Please read the instructions carefully before proceeding.

This exam has been designed to test your knowledge of the course that you did in term 1. The total number of marks you can achieve on this examination is 120. There are two parts. The first part comprises short answer questions, is worth 80 marks.

- You must complete EVERY question.
- You have 2 hours to complete this exam.
- Read all questions carefully.
- Attempt all questions.
- Some questions are multiple-choice: you are presented a range of possible answers and your task is to select the correct answer(s).
- For most of these questions, there is only one correct answer and you must only select one answer.
- For some questions, you will be required to choose more than one answer. These questions are clearly identified.
- Other questions are open response and you will be asked to write your answer in your own words.
- A statistical decision chart is provided on paper to help you answer some of the questions. These questions are clearly identified.
- Plain paper is also provided for doing calculations and making notes.

The paper materials MUST NOT be removed from the exam room.
Revising the examination

Summative Examination January 2014
Analysis Last Run: June 21, 2014 11:45 AM. Run Item Analysis again to display the latest question data

Test Summary

<table>
<thead>
<tr>
<th>Possible Points</th>
<th>Possible Questions</th>
<th>In Progress Attempts</th>
<th>Completed Attempts</th>
<th>Average Score</th>
<th>Average Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>121.0</td>
<td>36</td>
<td>0</td>
<td>162</td>
<td>58.78</td>
<td>01 hr 58 min</td>
</tr>
</tbody>
</table>

Discrimination

<table>
<thead>
<tr>
<th>Discrimination</th>
<th>Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 Good Questions</td>
<td>6 Easy Questions</td>
</tr>
<tr>
<td>16 Fair Questions</td>
<td>22 Medium Questions</td>
</tr>
<tr>
<td>3 Poor Questions</td>
<td></td>
</tr>
<tr>
<td>0 Cannot Calculate</td>
<td></td>
</tr>
</tbody>
</table>

Filter Questions

Select Question Type: All Question Types
Select Discrimination: All Discrimination
Select Difficulty: All Difficulty

<table>
<thead>
<tr>
<th>Question</th>
<th>Question Type</th>
<th>Discrimination</th>
<th>Difficulty</th>
<th>Graded Attempts</th>
<th>Average Score</th>
<th>Std Dev</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question 33: What are the assumptions that you need to bear in mind when choosing to calculate...</td>
<td>Multiple Answer</td>
<td>-0.01</td>
<td>68.52%</td>
<td>162</td>
<td>2.75</td>
<td>0.7</td>
<td>0.06</td>
</tr>
<tr>
<td>Question 7: Which of the following terms are considered to be more appropriate for describing...</td>
<td>Multiple Answer</td>
<td>-0.01</td>
<td>58.03%</td>
<td>162</td>
<td>1.17</td>
<td>0.58</td>
<td>0.05</td>
</tr>
<tr>
<td>Question 20: Read the following definition: Emerging theoretical considerations guide the...</td>
<td>Fill in Multiple Blanks</td>
<td>0.07</td>
<td>6.8%</td>
<td>162</td>
<td>0.14</td>
<td>0.51</td>
<td>0.04</td>
</tr>
</tbody>
</table>
Revising the examination

- **Automated analyses of:**
  - **Difficulty**
    - The percentage of students who got the item correct
      
      **Calculation:**
      \[ P = \frac{R}{T} \]
      where \( P \) is the item difficulty index, \( R \) is the number of correct responses and \( T \) is the total number of responses
  - **Discrimination**
    - An item’s ability to differentiate between high and low scorers on the examination
      
      **Calculation:**
      \[ P = \frac{R}{T} \]
      where \( P \) is the item difficulty index, \( R \) is the number of correct responses and \( T \) is the total number of responses
3. Research questions

- What are MA students’ attitudes towards and experiences of the delivery of high stakes computer-based exams?
- What factors predict MA students’ attitudes towards and experiences of the delivery of computer-based summative examinations?
### Exploratory research framework

**2013-14 cohort (n= 155)**
- Survey after summative exam (Jan 2014)
- Focus group (n=5) after summative exam, before release of results
- Qualitative content analysis on transcripts & free text survey comments:
- Focus on reception of computer-based testing methods (attitudes / experiences)

**2014-15 cohort (n=160)**
- Surveys after formative (mid-term) and summative exam (Jan 2015)
- Focus groups (n=18) after summative exam, before release of results
- Qualitative content analysis on transcripts & free text survey comments: repeated & combined with 2013-14 data to form rich picture of students’ reception of assessment methods
Participants (2014-2015)

- 28 Chinese females participated in the pre- and post- questionnaire
- Most were 20-24 years old
- 15 reported experience of e-assessment
- Most (21) reported using the internet everyday
- Some (15) reported pecking at the keyboard, while others reported touch typing (10)
- They were reasonably confident in their typing ability, but less so under timed conditions
## Most positive attitudes

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male students will perform better than female students in online exams</td>
<td>3.96</td>
<td>0.69</td>
</tr>
<tr>
<td>It is easier to change your answers in an online exam</td>
<td>3.89</td>
<td>0.79</td>
</tr>
<tr>
<td>Username and password login provide adequate security for online exams</td>
<td>3.71</td>
<td>0.71</td>
</tr>
<tr>
<td>Online assessment is appropriate for research methods</td>
<td>3.54</td>
<td>0.58</td>
</tr>
<tr>
<td>Marking is more accurate in online exams because computers don't suffer from human error</td>
<td>3.50</td>
<td>0.69</td>
</tr>
</tbody>
</table>
## Most negative attitudes

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I like to be able to highlight key terms when answering questions during an exam</td>
<td>2.00</td>
<td>0.77</td>
</tr>
<tr>
<td>It is easier to go back and check your answers in a paper-based exam</td>
<td>2.18</td>
<td>0.61</td>
</tr>
<tr>
<td>I get sore eyes if I have to read online for a long time</td>
<td>2.32</td>
<td>0.91</td>
</tr>
<tr>
<td>I would like to type essays in exams</td>
<td>2.36</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Engagement Issues: Socialisation of Learners

- **Issues**
  - Prior exposure to computer-based testing: rationale and perceived fairness & equity

"It’s kind of fair for most students because using a computer is almost a necessity for us and especially for our generation but it is not as much fair, as other generation."

"Some of our classmates after they have had some experience ... they return to school to get more experience in teaching. It could be some difficulty for them to use a computer in typing when they attend the examination, so it could take them longer time to get used to the system, so I think it could unfair for them."

---

Post-Test Focus Group 2014-15
Issues

- Keyboard proficiency under exam conditions: familiarisation with exam environment & controls

“I’m not used to using the keyboard because it’s different from laptop keyboard.”

“I feel in the real exam, I found there’s no correction tools for you to correct.”
Engagement Issues: Exam technique

- Online exam craft - question selection, time management...

“I do not like not having the ability to circle questions I am unsure about or make notes to myself about which questions to come back to. During written assessments, I often write all over my test questions with arrows, circles, and other brainstorming sketches and it is difficult to work through the online assessment without these techniques”

“In terms of time management...when we are doing the handwriting exam, I know what questions to I have, but in e_exam I just didn’t know what I am currently facing and I don’t know what kind of questions, you know closed or open question or is coming next ”

Post-Test Focus Group 2014-15
Engagement Issues: Revision strategy

- Management of self-study

“In China we will focus on the memory so we try to remember the long answers to these questions...... for the Chinese exam, I will remember all of the answers, long sentences, but I will not do this for this module.”

Post-Test Focus Group 2014-15
Engagement Issues: Assessment Design and Interface

**Issues**
- Organisation and presentation of question-set, preparation of user interface

"It’s no sense to put an open question for ten points at the beginning, so because our brain doesn’t work at the beginning to write/type so much."

*Post-Test Focus Group 2014-15*

"Random questions for each student don’t represent the level of difficulties, for some students could encounter long answer question at Q1 which gives little confidence of students to move on. More it could also waste time in trying to answer that question and therefore time is not enough."

*Questionnaire 2013-2014*
## Overall attitudes

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would like online exams to replace paper exams at university</td>
<td>2.71</td>
<td>1.05</td>
</tr>
<tr>
<td>I have received sufficient training to enable me to use the exam software</td>
<td>3.82</td>
<td>0.98</td>
</tr>
</tbody>
</table>

"Some students might not feel comfortable with the new method of testing but yes after they like they do it a lot of times, they will get used to it like the paper test, because the paper test is the beginning we do and now we want a little bit of change. Some students will find it hard to get the new things, so yes after they are doing this a lot of times, they will think it’s normal then people just accept it."

Focus Group Participant: 2014-15 Cohort
4. Reviewing our findings

Our research highlights the importance of:

- Socialisation of learners focusing on the aims and rationale for the computer-based assessment: key to acceptance of assessed learning task (Deutsch et al., 2012)

- Providing students with the opportunities for practice necessary to develop IT proficiency for computer-based testing and test-taking strategies for computer-based testing (Hillier, 2014; Zakrzewski & Steven, 2003)
  - A distinction should therefore be made between computer literacy (general computing experience) and using computers for formal and assessed learning tasks...
Our research highlights the importance of:

- Assessment interfaces should be flexible and intuitive to accommodate a range of test-taking strategies
  - ‘one-interface’ does not fit all students
- Equity in sequencing and order of questions (question value) to individuals
- Equity in the presentation of the exam environment across multiple venues (venue size & noise management; keyboarding quality)
The LEeAP framework

Socialisation
• Orientation of students to assessment methods

Preparation of students for assessment
• Guidance and preparation of students for computer-based testing:
  • Digital skills; Exam technique; Revision strategy

Assessment design and interface
• Organisation and presentation of question-set, preparation of user interface and assessment venue:
  • Assessment design; Design of assessment interface; Preparation & management of assessment centre(s)
5. Impact of findings on our practice (2015-16)

Socialisation
- Briefing on general exam regulations
- Video tutorial of exam environment controls
- Examples of each question type on VLE training module

Preparation of students for assessment
- Formative assessment aligned with summative (question types, difficulty and time pressure)

Assessment design and interface
- Questions blocked according to marks available
To navigate the questions in the exam you can use the blue bar above where the questions are displayed.

ONLY when you have finished:

1. Go to the last question
2. Click Save and Submit
3. Click OK

Do not click SAVE AND SUBMIT if you have not finished the exam. You will NOT be able to re-enter the exam.
Exam design

- Attention to proportion of short answer (essay) questions to MCQs
- Use of blocks (question type / difficulty)
- Randomisation by question type (higher value questions served up first to all)
How transferable is the LEeAP framework to your PGT students?

To what extent is this framework relevant only to ‘greenfield’ institutional contexts?

Are we dealing with a transient phase with international PGTs, where they acquire the skills to tackle computer-based assessments? Is it likely that future cohorts will arrive with first-hand experience of high stakes computer-based testing?
Further information

- **VLE Exam video case study:**

- **Working paper:**
  - An earlier version is available at: [http://tinyurl.com/LEeAP](http://tinyurl.com/LEeAP)
References


Thank you!

Richard Walker
E-Learning Development
Zoe Handley
Department of Education

University of York, UK
Session feedback:

With thanks from your hosts

Professor Geoff Crisp,
PVC Education, University of New South Wales
g.crisp[at]unsw.edu.au

Dr Mathew Hillier,
Office of the Vice-Provost Learning & Teaching
Monash University
mathew.hillier[at]monash.edu

Recording available
http://transformingassessment.com