Report on
Summative e-Assessment Quality

Report on Summative E-Assessment Quality (REAQ)

Prepared by
Lester Gilbert, Veronica Gale, Bill Warburton, and Gary Wills
of the Learning Societies Lab
School of Electronics and Computing Science
University of Southampton

Final Report, May 2009

Executive Summary

Commissioned by the Joint Information Systems Committee (JISC) in 2008, the 'Report on Summative e-Assessment Quality (REAQ)' project surveyed quality assurance (QA) activities commonly undertaken in summative e-assessment by UK Higher Education (HE) practitioners and others. The project focused on what denotes high quality in summative e-assessment for the interviewees and the steps that they take to meet their own standards. An expert panel guided the project.

What denotes high quality summative e-assessment

Expert opinion focused, in this order of priority, on:

- Psychometrics (reliability, validity),
- Pedagogy (mapping to intended learning outcomes), and
- Practical issues (security, accessibility).

What 'high quality' meant to our interviewees depended on the role they played in the process of creating and using e-assessments. They listed the following matters, in this order of volume:

- Using the medium to give an extra dimension to assessment, including creating e-assessments that are authentic to the skills being tested;
- Issues around delivery including security, infrastructure reliability, and accessibility;
- Fairness and ease of use;
- Supporting academic, managerial, and organisational goals;
- Addressing the intended learning outcomes; and
- Validity and reliability, mainly in their 'non-psychometric' senses. Interviewees with the role of learning technologist (or similar roles designed to aid academics in the use of e-assessment) used these terms in their psychometric senses.

Interviewees focused on the e-assessment issues that were foremost in their mind. As processes to deliver e-assessment are rarely embedded in institutions at present, interviewees described spending time and effort on practical issues ensuring that e-assessments would work effectively.

Many of the quality characteristics identified by the interviewees as important in summative e-assessment are measured by psychometrics. Although some academics use these measures, the report suggests that more could benefit from using psychometric evaluation.

Steps needed to produce high quality e-assessment

Expert opinion focused on:

- Establishing sets of steps to follow for both content and quality management;
- Identifying, using, and developing relevant standards for both content and quality management;
- Identifying metrics for both content and process; and
- Capability maturity modelling as an encapsulation of these three essential elements of a quality management process.

Interviewee comments fell under a variety of rules of thumb or suggestions for useful steps, such as: noting that the effort needed to write e-assessments, their marking

schemes, and to construct feedback is front-loaded; starting with easier questions and making later questions more difficult; checking assessments with subject matter experts and high performers; identifying 'weak' questions and improving or eliminating them; reviewing question content to ensure syllabus coverage; getting help for academics who usually have very limited knowledge of psychometrics; attending to security; and using accessibility guidelines. In summary:

- Heuristic steps for both content and quality management, and
- Accessibility standards.

Many interviewees assumed that e-assessments were:

- Valid if they were created by the academics responsible for the course, and
- Subject to the same quality assurance processes as traditional assessments as well as those required specifically for e-assessment.

The report questions these assumptions.

Recommendations

The report makes a number of recommendations to support academics creating high quality summative e-assessments, including:

- A toolkit for the end-to-end process of creating e-assessment should be developed.
- A practical guide to the steps involved in creating and maintaining an e-assessment system.
- Guidelines for the quality assurance of e-assessments.
- Psychometric measures for assessing the quality of item banks rather than individual questions, for assessing, tracking, and reporting the quality of banked items during their lifecycle of use.
- Development and extension of existing psychometric theory to include multi-staged and optional stepped constructed response questions.
- Workshops and support materials to disseminate good practice in the use of psychometrics for selected response items and for questions employing constructed responses.
- Workshops and support materials to disseminate good practice in question creation and meeting educational needs beyond simple selected response items, possibly subject based.
- Accessibility and user interface guidelines for deploying e-assessment, in particular addressing the use of browsers.
- Guidelines for the use and role of MathML for expression recognition in e-assessments.
- A repository of exemplars of good practice for both selected response and constructed response questions.
- JISC and other community calls for and sponsorship of e-assessment bids should consider where and how bidders should incorporate appropriate psychometric measures in their proposals.
- Commercial vendors should improve the accessibility of their psychometric reports to all stakeholders, possibly simplifying them to encourage take-up of their contents.

Contents

Introduction

Report process

This report has been commissioned by the Joint Information Systems Committee (JISC). The project undertook a survey of what quality assurance (QA) activities are commonly undertaken by summative e-assessment practitioners. The survey focused on what denotes high quality in summative e-assessment for the interviewees and the steps that they take to meet their own standards.

Interviewees were chosen as representatives of:

- current practice in Higher Education (HE) contexts both in the UK and overseas;·and
- current practice in professional, commercial, and industrial contexts.

Part 1 of the report looks at the meaning of 'high quality' in terms of the products or content of summative e-assessment. It quotes the characteristics that interviewees told us are important, summarizes the theoretical framework and international good practice, and then compares the two. The comparison section is where the experts discuss the points that they believe need to be addressed. The key points from this part of the report are summarized in the section on tips and traps.

Part 2 of the report looks at the meaning of 'high quality' in terms of the processes of creating, delivering, and evaluating summative e-assessment. As for Part 1, it quotes the characteristics that interviewees told us are important, summarizes the theoretical framework and international good practice, and then compares the two. The key points are summarized in the section on tips and traps.

In Part 3 we make recommendations for readers of this report, including summative e-assessment practitioners and JISC.

This process is illustrated in Figure 1.

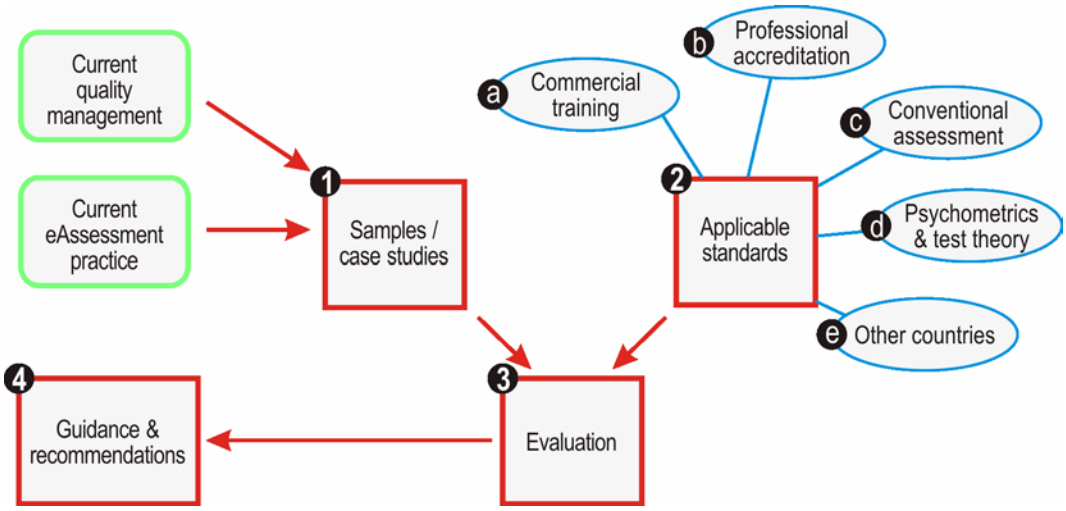

Figure 1: Project components and activities

Expected components of e-assessment in UK HE

What 'high quality' means to you depends on your role in the process of creating and using e-assessment. The major stakeholders are illustrated in Figure 2, and their views are reflected in the comments of the various interviewees in Parts 1 and 2.

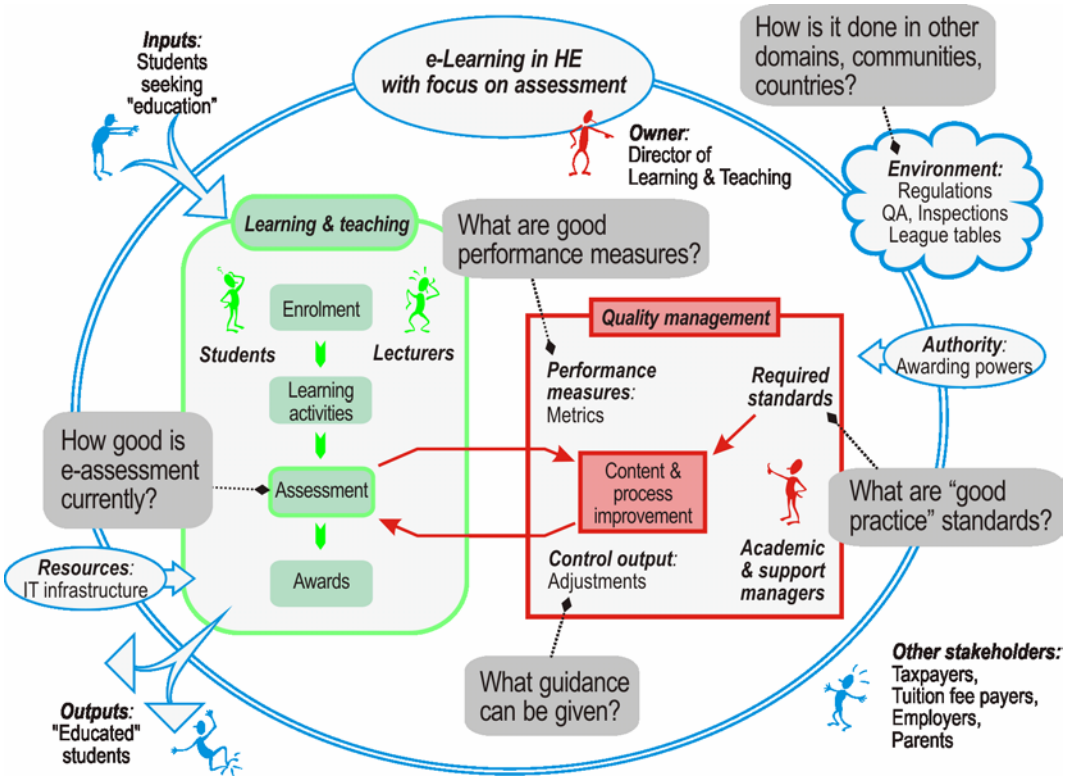

Figure 2: Systems diagram of e-assessment within HE

Figure 2 also illustrates the questions (grey) the project sought to answer around summative e-assessment quality in UK HE. The question of the quality of e-assessments themselves is illustrated within the learning and teaching process of the figure (green) and is addressed in Part 1 of the report, while questions around the quality management of e-assessments are illustrated within those processes of the figure (red) and are addressed in Part 2 of the report.

JISC/QCA definition of e-assessment

This project used the JISC/QCA definition[1] for e-assessment:

> "E-assessment is the end-to-end electronic assessment processes where ICT is used for the presentation of assessment activity, and the recording of responses. This includes the end-to-end assessment process from the perspective of learners, tutors, learning establishments, awarding bodies and regulators, and the general public."

[1] JISC e-Assessment - An overview of JISC activities, Version 2: July 2008. Retrieved March 2009 from http://www.jisc.ac.uk/whatwedo/themes/elearning/assessment.aspx

We therefore excluded assessment tools that only handle electronic submission or marking.

Examples of e-assessment that interviewees gave us

Many of the examples we were given in response to the question, 'Please give us examples of good e-assessment' are taken from practices in Higher Education in which each question or item is either all right or all wrong, i.e., one-mark items where the answer is computer-marked and not subject to human interpretation. These are known as objective items and they are typified by a multiple choice question format or one of its several variants (drag and drop, hotspot, etc). Such selective objective practices have a well-developed theoretical framework produced over many years in classical test theory (CTT) or item response theory (IRT). However, this report has sought to identify all practices of summative e-assessment in UK Higher Education and identify the quality issues associated with the delivery of all strands of such activity. So, this report also includes the widespread use of e-assessment of questions such as those that require a constructed mathematical expression as an answer. These practices have been employed in a variety of universities over many years through projects like CALM[2], Mathwise[3] , SCHOLAR[4] and MathsAssess[5]. Typically, in this approach, questions have several parts to be answered and in some cases steps can be accessed optionally to guide students to these part-answers.

Acknowledgements

We wish to thank the following Expert Panel for their guidance and support during this project. Please note that these experts have provided guidance and support during the project but do not necessarily endorse every recommendation and view expressed in this report. The Expert Panel:

Cliff Beevers, Heriot-Watt University
Paul Booth, Question Tools
John Kleeman, Question*mark*
Harvey Mellar, London Knowledge Lab, Institute of Education
Greg Pope, Question*mark*
Chris Ricketts, Peninsula College of Medicine and Dentistry
Denise Whitelock, The Open University, UK

We also thank the following contributors for their help in identifying what high quality e-assessment means in practice and how to achieve it.

Helen Ashton, Heriot-Watt University
Les Carr, University of Southampton
Hugh Davis, University of Southampton
Silvester Draaijer, Vrije Universiteit Amsterdam
E. A. Draffan, University of Southampton
Bobby Elliot, Scottish Qualifications Authority
Bill Foster, University of Newcastle
Alex Furr, University of Southampton
Joan Gavin, University of Plymouth

[2] http://www.calm.hw.ac.uk/
[3] http://www.bham.ac.uk/mathwise
[4] http://scholar.hw.ac.uk/
[5] http:// www.qtitools.org

Robert Harding, Cambridge Assessment
Jenny Hayes, Question Tools
Phil John, Heriot-Watt University
Sally Jordan, The Open University, UK
Helena Knowles, University of Southampton
Kenji Lamb, JISC, N&E Scotland
Jackie Lupton, University of Southampton
Nora Mogey, University of Edinburgh
Ben Palmer, The Open University, UK
Priska Schoenborn, University of Plymouth
Karl Wall, Institute of Education, University of London
Martin Youngson, Heriot-Watt University

Method

We interviewed 23 people, most individually and eight as a group, using a standard list of questions (see Appendix J).

1. What 'high quality' means

Interviewees were asked, 'What denotes high quality in summative e-assessment?' Section 1.1 of this part focuses on what interviewees told us. Section 1.2 outlines the theoretical and international practice for high quality. Then section 1.3 compares what practitioners told us with what the theoretical and international practice suggests. The final section in this part, 1.4, pulls the threads together by summarising some tips and traps from the practitioner and expert points of view.

Our expectations are illustrated in Figure 3. We anticipated we would hear about delivery issues, and about item and test quality issues in terms of relevant psychometric measures and pedagogic practice.

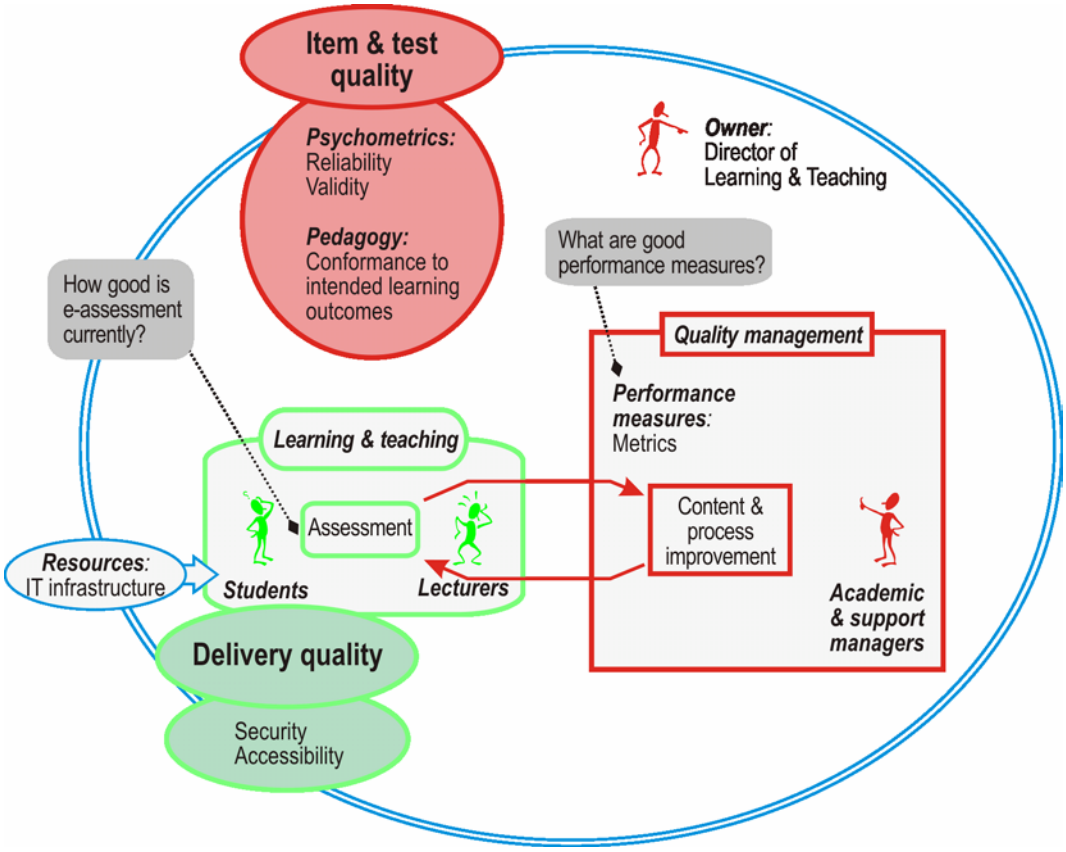

Figure 3: Project expectations of Part 1 issues

1.1 What interviewees told us

All interviewees gave us a list of characteristics that together denote 'high quality' summative e-assessment from their perspective. Each list was different, although interviewees identified some characteristics in common. There is therefore no generally-accepted definition of high quality for summative e-assessment in UK higher education (HE) and the responses we received show it is not a simple matter to define it. As one interviewee said,

> "Every aspect has to be right as e-assessment will fail at its weakest link. There is a vast list of components that must be done properly from setting clear and

> defined questions that are relevant to the syllabus through to equality of marking." (Robert Harding, Cambridge Assessment)

We have identified the first response that interviewees gave us as this shows the characteristics that have a top priority from their point of view. We also identify the second and subsequent characteristics that interviewees gave us. As one of our expert panel put it,

> "Quality is in the eye of the beholder." (Paul Booth, Managing Director, Question Tools)

Security and fairness

Of the 23 people we interviewed, 13 highlighted security and 14 emphasised fairness as important for 'what denotes high quality summative e-assessment' from their point of view. These comments were across the different job roles and work contexts of the interviewees.

From the providers' point of view (i.e., HE institutions, commercial and professional organisations that author and deliver e-assessment) and that of certain users of the assessment results (such as senior managers within the provider, other HE institutions and employers) the issue of security focused on, 'Is the individual answering the questions who they say they are?' To ensure high quality security, interviewees mentioned a variety of techniques such as requiring students to show identity cards in an invigilated assessment centre, and asking students to sign an electronic declaration that they are who they claim to be. These issues are explored further in the processes part of the report (Part 2: Processes that produce high quality summative e-assessment).

Fairness was considered a key characteristic of high quality e-assessment for providers, users of results and students. From the providers' point of view this meant making e-assessments accessible to all students being assessed and ensuring that individual and stepped questions:

> "Give sufficient context in the question to make it a fair assessment of student's skills." (Karl Wall, Institute of Education, University of London)

From the student perspective, interviewees told us that clear and unambiguous questions are most important.

Many of the factors that were identified as characteristic of high quality e-assessment apply to all types of assessment. Other factors focus on what is important for e-assessment in particular. We have organised the comments so that the ones that apply specifically to e-assessment are placed first, as these may be of particular interest to readers. These comments fall into the following categories: issues around delivery: ease of use; creating assessments that are appropriate to the medium and using the medium to give an extra dimension to the assessment. Characteristics our interviewees identified as important for high quality e-assessment that also apply to other types of assessment are: addressing the intended learning outcomes effectively; validity and reliability; and providing information that supports academic, managerial and organisational goals.

The following quotes are organised by interviewee role. After each quote, a reference is provided in parentheses to the source.

Role: Course chair or subject leader

The people we interviewed either perform these roles themselves or work closely with the people who take these roles in their organisation and can therefore identify what high quality means to them.

Effective delivery

We asked 'What denotes high quality in summative e-assessment?' Of the 14 interviewees in these roles, nine identified delivery issues of infrastructure and support in their immediate response.

> "Infrastructure and support for e-assessment creation and delivery is vital." (Cliff Beevers, Heriot-Watt University)

> "One important part of the implementation of e-assessment was when the utilisation of e-assessment in these modules became school policy. These assessments were moved from 'belonging to the lecturer' to 'belonging to the school'. [Now] there is school administrative and technical support with positive encouragement from the Head of Department." (Bill Foster, University of Newcastle)

The summative assessment for one course discussed with us has been changed from e-assessment back to a paper-based examination. A key driver for the change was that the school's course metrics did not allocate sufficient time to conduct an e-assessment appropriately.

Another interviewee identified:

> "One of the biggest problems with e-assessment is having the resource to deliver the e-assessment to 1000 students at a time." (Chris Ricketts, Peninsula College of Medicine and Dentistry)

Uses medium to give an extra dimension to assessment

As a second or subsequent response to our question on high quality, 11 interviewees said that it was when the medium was used to give an extra dimension to the assessment that could not be achieved with more traditional methods.

As one interviewee stated:

> "If you are going to do the exam online you need to do so for good reasons, i.e., that e-assessment offers a clear advantage over traditional exams" (Helena Knowles, University of Southampton)

Authentic

All 11 of these interviewees mentioned that the medium of e-assessment offers opportunities to present assessments in ways that offer more contextual detail and are therefore more authentic to the skills being assessed.

> "E-assessment facilitates more authentic assessment [than traditional methods] through multimedia, simulations etc" (Bobby Elliot, Scottish Qualifications Authority)

Interviewees explained how they used the 'e' medium to create assessments that are more authentic to the specific subject skills being tested. For example:

> "The i-assess product was devised to accept mathematical expressions as answers and provide the possibility of a stepped and staged approach to mathematical questions." (Bill Foster)

Using 'e' in this way gives an extra dimension to the assessment that would be much more time-consuming to provide or could not be provided via traditional examinations. Many of the examples where 'e' adds a dimension missing in other media come from medicine, nursing, dentistry, and psychology:

> "The unique ability of the medium to provide a virtual patient" (Denise Whitelock, The Open University, UK)

However:

> "The clever bit is getting the right balance between the richness of the question, for example by showing qualitative information, and the objective information needed to focus the able student on the anticipated correct answer". (Helena Knowles)

Assessing a range of skills levels

Interviewees also gave us examples of how the 'e' medium enabled them to gauge if students could find the correct answer with some help. Hints and partial credits are being used in mathematical subjects and:

> "In modern languages we have used an e-assessment approach that provides hints for students doing translations, for example on the tense of a verb." (Nora Mogey, University of Edinburgh)

> "In low-level courses students are given up to three attempts to pass the summative e-assessment. If their first answer is wrong, they are told immediately and given another attempt. The number of marks they earn is reduced in line with the number of attempts they need to get the answer correct." (Ben Palmer, Head of Assessment, Credit and Awards, The Open University, UK)

Assessing extra facets of skill

Another example of the extra dimension added by the 'e' medium was asking students to specify how confident they felt in each of their answers. This enabled the tutors to assess a student's level of confidence in their skills:

> "an essential characteristic of a successful student for the Applied Medical Knowledge Progress Test on medicine." (Chris Ricketts).

Providing feedback

One interviewee's first response to our question about high quality e-assessment was:

> "Accurate marking and helpful feedback." She went onto explain, "Personalised and quick feedback is important for students to be able to develop their understanding. The medium of e-assessment was chosen because we wanted quality feedback to be provided quickly for both formative and summative tests. Students are at a distance and however hard tutors work, they cannot provide feedback instantaneously." (Sally Jordan, The Open University, UK)

The group of eight academics also identified providing feedback as an indication of quality in summative e-assessment:

"Students need to get their scores quickly and to be told about their mistakes so they can learn." (Group discussion)

Example of an objective test with feedback

The Peninsula Medical School website for students introduces the Applied Medical Knowledge Progress Test. It explains that the test:

"is one of the key features of the School's approach to assessment. The Progress Test, which is delivered in a multiple-choice question format, is designed to assess long-term and functional knowledge rather than detailed and easily forgotten 'facts'. It is a measure of how much you are learning, not how good you are at revision, cramming or rote memorisation. Following every test that you take, four per year in total, you will receive your grade and percentage score as well as the mean percentage of each test. You will receive a line graph that plots your progress against the mean cohort progress over time."

Sample Applied Medical Knowledge Progress Test Question[6]

A 75 year old female, presents to the emergency department in a state of acute respiratory distress. From her accompanying husband, Phil, you ascertain that she has suffered with chronic allergic asthma since childhood. She has brought both of her inhalers with her, and is carrying them in her right hand.

Tests reveal a heart rate of 115 beats/min, systolic blood pressure of 95 mmHg. Pulse oximetry reveals that oxygen saturation is 88%. PEF is 47%, and there is evidence of cyanosis of the lips.

Following initial stabilisation, the duty registrar suggests that you cannulate the patient to obtain arterial blood gas samples.

What would be the most appropriate location for this procedure?

A) Right femoral artery

B) Left tibial artery

C) Left radial artery

D) Right brachial artery

E) Brachiocephalic trunk

F) Don't know

Feedback

A) Right femoral artery

Incorrect. The femoral artery is used in emergencies for arterial blood gas sampling if an accessory artery cannot be accessed. However, its position as the major artery supplying the lower limb mean that any complications in the procedure may compromise the entire blood supply to that limb. In elderly patients especially, dislodgement of atheroma from the femoral artery may occur.

B) Left tibial artery

Incorrect. The tibial artery's position as the major artery supplying the distal lower

[6] http://www.pms.ac.uk/pms/undergraduate/assessment.php

limb mean that any complications in the procedure may compromise the entire blood supply to that part of the lower limb. It is also inaccessible for palpation and stabilisation.

C) Left radial artery.

Correct. The radial artery has a good collateral supply from the ulnar artery, is near to the surface and is relatively easy to palpate and stabilise. The chance of atheroma dislodgement is small. As the patient is carrying items in her right hand, there is a clue that she is right hand dominant. The correct side for sampling would be the non-dominant side.

D) Right brachial artery

Incorrect. Although the brachial artery may be used for this procedure, it does not have the greatest combination of accessibility, risk minimisation and collateral supply. It may be used if other sites are not available. As the patient is carrying items in her right hand, there is a clue that she is right hand dominant. The correct side for sampling would be the non-dominant side.

E) Brachiocephalic trunk.

Incorrect. The brachiocephalic trunk is not used for this procedure due to its inaccessibility and potential for complication.

F) Don't know

The radial artery has a good collateral supply from the ulnar artery, is near to the surface and is relatively easy to palpate and stabilise. The chance of atheroma dislodgement is small. As the patient is carrying items in her right hand, there is a clue that she is right hand dominant. The correct side for sampling would be the non-dominant side.

References

Ballinger A and Patchett S (2004). *Saunders' Pocket Essentials of Clinical Medicine (3rd ed).* Edinburgh: Saunders, p737-739.

Academics are also using e-assessment to manage their assessment strategy in ways that it would be very difficult to achieve with more traditional methods.

> "We can make assessments available on demand. Students attempt when ready." (Helen Ashton, Heriot-Watt University)

This approach enables students to take greater responsibility for their assessments and to manage their study time more effectively.

Examples

Course: First year maths and statistics degree courses at Newcastle University

Institution: University of Newcastle

Description: Randomized components in an equation so the same 'question' is not used twice.

(Bill Foster)

Course: M.Sc. E-Learning

Institution: University of Edinburgh

Description: A wiki is used for summative assessment. It contributes 50% to the final mark and illustrates online collaborative assessment. All students work on the wiki and each student is assessed separately using defined criteria. The final product was **very** impressive – a huge wiki created by 20 students.

(Bobby Elliot)

Essay Exams on computer

Institution: University of Edinburgh

Description: We are undertaking research around, and implementing, students sitting traditional essay exams but where they type responses on their own laptops rather than handwriting. This is felt to be more consistent with their approach to the essay writing they do during their learning – many students may never have constructed an essay on paper prior to taking their exam. Nothing has changed within the process apart from the students typing rather than writing – students still sit in an invigilated room, for a timed exam, get the question presented on paper and (currently) submissions are still printed for marking. A substantial amount of research has been done to ensure this is appropriate, fair (we are considering offering students the choice of typing or writing) and technically practical. It requires good security, but there is potential to use the system for open internet exams.

(Nora Mogey)

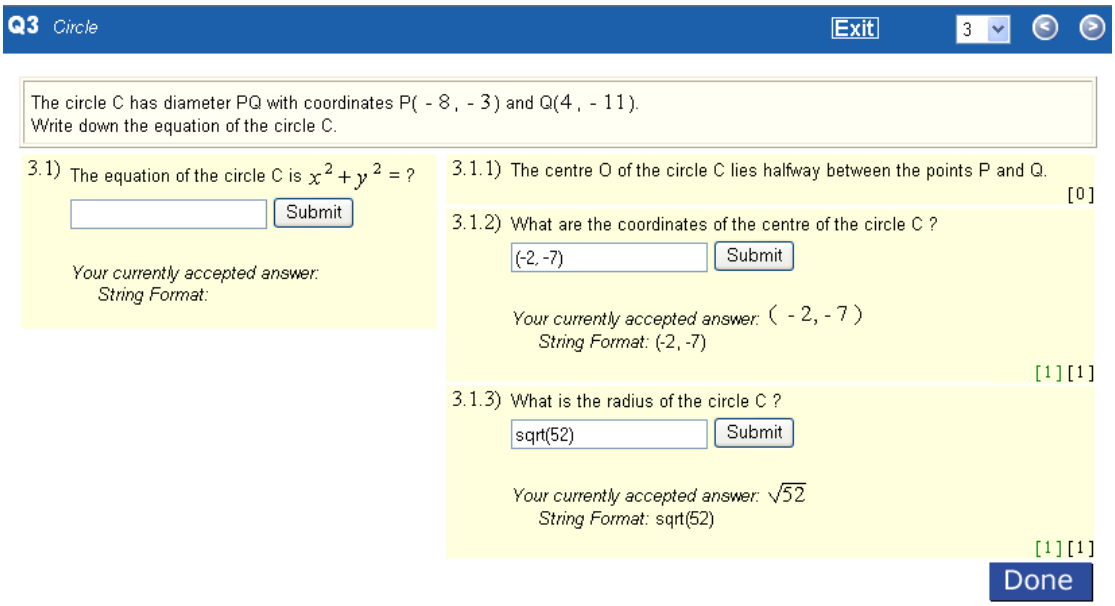

Figure 4: A summative question showing a student using steps (Cliff Beevers)

Please see Appendix K for a discussion of this type of question.

Addresses ILOs

The immediate response to our question from three of the interviewees in this category was “addressing intended learning outcomes”. For example:

> “Assesses intended learning outcomes, tests at appropriate level.” (Helena Knowles)
>
> “Appropriate alignment to what you are testing” (Chris Ricketts)

Another person specified that high quality summative e-assessment should be “fit for purpose” which he explained as demonstrating that the student has met the intended learning outcomes of the course.

Valid and reliable

The group of eight experienced academics provided the following list of characteristics after first identifying ‘infrastructure and support’ as vital to high quality summative e-assessment:

> “Validity, reliability, predictability, usability, accessibility, appropriate, challenging, authentic and engaging” (Group discussion)

Another person identified reliability and validity as important in their second or subsequent response to the question.

One interviewee told us that reliability was important and explained that this meant ensuring the computer system worked well. Other interviewees mentioned characteristics that could be interpreted as valid and reliable such as checking that questions discriminated between more able and less able students. We discuss this in more detail in Part 2.

Role: A content manager in a commercial organisation assessing competence

Effective delivery

We spoke to a company that supplies e-assessment software tools about the requirements of their commercial clients to get a different perspective on our question, ‘What denotes high quality in e-assessment’. Again, the needs of their clients depend on the role that the individual takes in the creation and use of e-assessment.

From the content manager’s point of view delivery issues are a priority:

> “A large test bank with lots of validated questions so that candidates get a very different test each time they take it.” (Jenny Hayes, Project Director, Question Tools)
>
> “Limits on number of attempts should be “soft” not “hard”. A hard limit of one attempt for example will always be met at some stage with a good reason for an exception (e.g. fire alarm, illness). They need to be sure that exceptions are possible. For example, in the QT system candidates can take an exam, but to take it again they require a member of staff to enter their username and password. There is a clear message explaining that the exam has recently been attempted by the candidate, and that the member of staff is taking

responsibility for approving another attempt, and is being recorded as the person who approved the retest." (Paul Booth)

Uses medium to give an extra dimension to assessment

For commercial managers using e-assessment to prove competence, authenticity is a key quality characteristic:

"The visual possibilities of e-assessment can be used to make the test as close to real life as possible." (Jenny Hayes)

For example, Figure 5, railway workers are asked to select the correct tensioner:

Figure 5: Example of a MCQ using photographs

Using a photo of the actual end assemblies available on the job brings the question close to the reality of the task. It also means that the context of the question is clear. The presentation of the question is also straightforward, making it easy for the candidate to understand what needs to be done to answer it.

Another example, Figure 6, shows the use of a diagram and drop-down menus to select the correct responses:

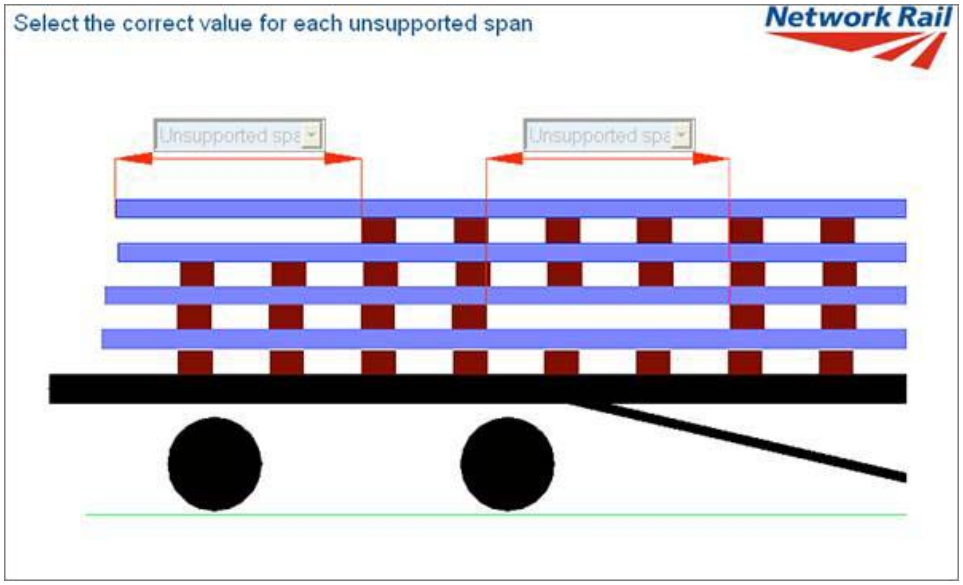

Figure 6: Example of a menu question

This question uses a standard diagram that shows possible loading patterns on a wagon and asks the candidate to select from a range of measurement options. It is asking a real-life question about how loads can be stacked safely in a way that would be familiar to the candidate. This question focuses clearly on the key job skill to be tested.

The following example, Figure 7, uses drag and drop to simulate real job decisions:

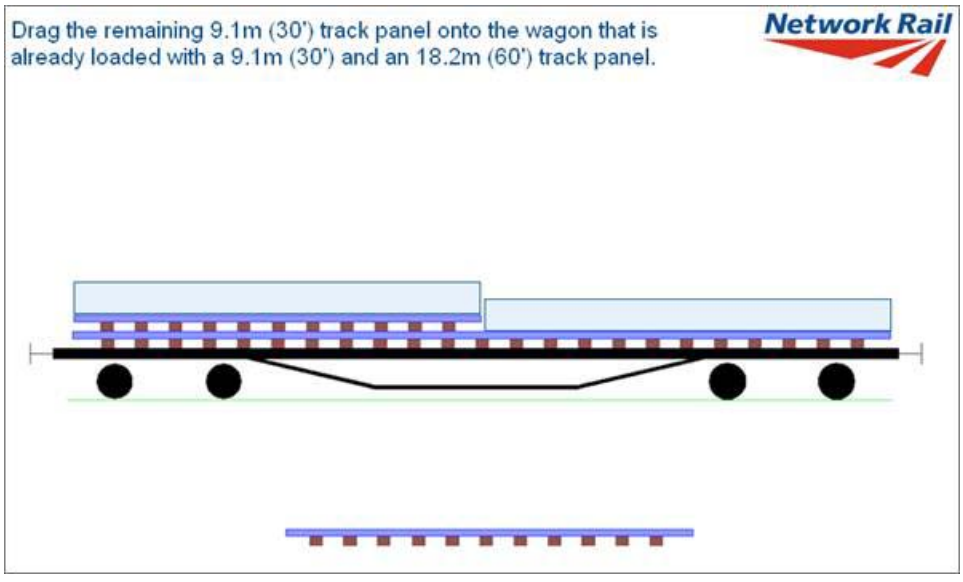

Figure 7: Example of a drag and drop question

This question asks the candidate to drag a load onto a wagon – giving them two options for safe positioning. This question reflects real-life loading problems and the

act of dragging the load onto the top of the existing load is similar in concept to the way the load would be lifted and placed in real-life.

> "What is also good about these questions is that they offer a variety of interactions, so the candidate is more likely to stop and think about each question individually than to 'bang through' a series of multiple-choice questions." (Jenny Hayes)

While the question above is an on-screen simulation of what happens in real-life, it is of necessity limited by the reality of the nature of the task. Where tasks are screen-based, it is possible to set questions that simulate the actions actually needed in the real task (see below).

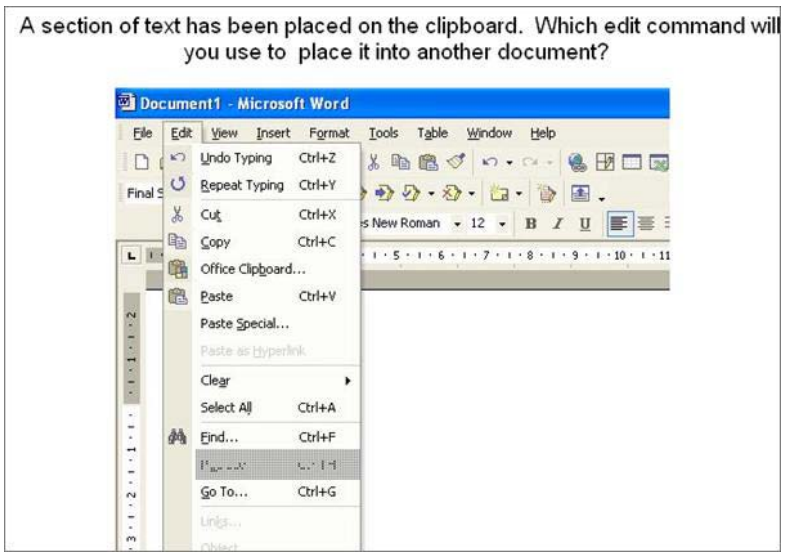

Figure 8: Example of a simulation question

In this question, Figure 8, the candidate needs to click on the appropriate command on the menu that is dropped down on the screenshot – exactly as you would if you were trying to paste a piece of text in a document.

When asked about other methods, our interviewee commented:

> "At the scale we have to operate, automated marking of free text is in its infancy, is unreliable and is best avoided." (Ben Palmer)

Another example in a commercial context of how 'e' can add an extra dimension uses adaptive test:

> "Manpower uses an adaptive e-assessment that measures English use for speakers of other languages. This test was developed by 'Cambridge ESOL'; it focuses on vocabulary, sentence completion and comprehension and involves candidates typing responses. It gives a very reliable result in under an hour." (Robert Harding)

Supports academic, managerial and organizational goals

For commercial managers, high quality e-assessment means that they can rely on the results generated by the test:

> "If a person passes an e-assessment s/he should have the essential knowledge for performing the relevant tasks in his/her job for which s/he holds a competency. In other words, if a person passes the Handsignaller test, then s/he is able to answer questions on the underpinning knowledge needed by an active Handsignaller." (Jenny Hayes)

To enable managers to make decisions based on the results of the e-assessment the test as a whole, not just the individual questions, needs to reflect the reality of the task. Our interviewee explained:

> "Every time an e-assessment is taken, it should ask questions on the core knowledge needed to reach competence. Questions on useful but not essential knowledge should be labelled as optional and weighted accordingly." (Jenny Hayes)

Role: An author

Effective delivery and easy to use

Thirteen of the interviewees told us about the authors they work with to create e-assessment. From the author's point of view, high quality means:

> "An e-assessment tool that is reliable (i.e., it won't crash and will produce data that is trustworthy) and is easy to use to construct the e-assessment. This can be a problem as authors tend to only use the tool once a year." (Alex Furr, University of Southampton)

Twelve of the interviewees mentioned that their institution employs learning technologists or administrative colleagues (who are trained in the use of the tool) to assist authors to create and/or input questions.

Role: A learning technologist, computer system expert or academic adviser

Valid, reliable, accessible

During this survey, the first response interviewees gave to the question on 'What denotes high quality in e-assessment?' was rarely that it should be valid and reliable as these terms are used in psychometrics. However, two of the five interviewees who fall into this role category used these terms in their initial response to the question.

> "High quality summative e-assessment means it is valid, reliable and accessible. It should comply with classical test theory." (Silvester Draaijer, Vrije Universiteit Amsterdam)

Another interviewee's first response was to think about what type of test is deliverable but followed this observation with:

> "Also needs to be valid, reliable and therefore tested". (Joan Gavin, University of Plymouth)

Supports academic, managerial and organizational goals

One interviewee's initial response to our question on high quality was that e-assessment should be:

> "Predictable, accurate, cost-efficient, demonstrably valid, auditable, legally-defensible, equitable (hard to cheat and doesn't discriminate against students with disabilities), satisfies stakeholders." (John Kleeman, Question*mark*)

This reflects the importance of e-assessment giving usable results. John Kleeman's colleague pointed out that:

> "In the US students sometimes sue organisations over their results" (Greg Pope, Question*mark*)

Therefore high quality e-assessment must be legally defensible.

Example

Psychometric practice in USA

"I would say essential things that should/do get done are:

- Questions should be thoroughly reviewed after they are initially tested to ensure that they do not have obviously psychometric flaws (e.g., negative CTT discrimination statistics). If questions are found to be flawed they should be dropped from the test and participants should have adjusted scores produced.

- Test level psychometric information should be evaluated especially the internal consistency reliability (e.g., Cronbach's Alpha)." (Greg Pope)

Easy for students to use

The interviewees in this category stressed that the medium should not disadvantage students. These comments were made as their second or subsequent response to our question on high quality e-assessment.

> "User-friendliness is important such as being able to revisit questions to change answers before submitting the assessment. It should be obvious how to use the assessment. Content and language should be clear and logical." (Priska Schoenborn, University of Plymouth)

> "Students are bothered by technical problems and poorly-designed screens. They are used to seeing computer games with high quality visuals." (Silvester Draaijer)

Appropriate to medium

> "In a first year undergraduate course we use simple multiple-choice questions to test the students' level of knowledge. By the third year of the course we are using e-assessment to test underpinning knowledge. The questions are created very carefully." (Joan Gavin)

> "The question types in [our e-assessment tool] are limited, so we focus on asking questions at a lower level. We need better tools to ask questions that require more discursive answers because of the variety of good responses. The

exception is task-based assessment where more complex skills can be addressed using simple question types." (Bill Warburton, University of Southampton)

Uses medium to give an extra dimension to assessment

High quality summative e-assessment means:

> "Making use of the possibilities of e-assessment, such as using images, multimedia and different question types. It is possible to make e-assessment more engaging than traditional methods. There is a broader sense of what assessment can be when you use the possibilities offered by e-assessment, such as digital portfolios, 360 degree feedback, plagiarism detection and peer reviews as well as the Question*mark* tests." (Silvester Draaijer)

A further four interviewees in this category told us as their second or subsequent response to our question on what denotes high quality that summative assessment:

> "Should promote learning, so feedback (beyond just a score) is important. The 'e' medium gives authors an opportunity to provide tailored student feedback for comparatively less effort than with traditional assessments." (Karl Wall)

> "Feedback should be provided in summative e-assessments, including feedback for the correct answer. Feedback should explain how to derive the correct answer and give further resources. The system should make students look at their feedback before they get their score." (Priska Schoenborn)

Addresses ILOs

One interviewee's first response to our question was that e-assessment should be "fit for purpose" and explained that this means high quality summative e-assessment:

> "Assesses the expectations for what students will be able to demonstrate having completed the course successfully. This skill relates to the learning outcomes for the course. But it is usually at a much more detailed level when setting an individual question. You must decide to what extent you fragment the knowledge and skill to create workable e-assessment questions. High quality e-assessment also assesses the appropriate level of skill. For example, reproducing information, applying a process to a novel problem or extrapolating from a specific example to a principle. And it reflects the way in which the students will interpret the question or find answering the questions most appropriate, based on their life experiences. For example, ask for two numbers that add up to equal 4 and you will get a very different range of potential correct answers from 9 year olds and maths graduates." (Karl Wall)

Another interviewee identified meeting intended learning outcomes as her second or subsequent response to our question.

Role: A deliverer

One interviewee was responsible for delivering e-assessments in her HE institution. Her immediate response to our question on high quality was that:

> "We need time to develop an e-assessment policy. At the moment a small percentage of summative assessments are 'e' and they take a disproportionate

amount of time to prepare and deliver. For example, there are many practical issues around receiving the test in time to check it works correctly on the system, and invigilators need extra training." (Jackie Lupton, University of Southampton)

Explaining how recent the use of e-assessment is for the university, a computer systems expert said:

"Up until 2003, the University's Examination regulations actually prohibited the use of e-assessment for delivering exams - they had a clause saying that marks could not be transferred electronically, presumably to minimise the risk of leakage." (Bill Warburton)

Role: A student

Effective delivery

Interviewees told us that students like:

"E-assessment in general, unless they perceive they have been disadvantaged." (Joan Gavin)

"Short tests with varied question types, not just multiple choice. But don't have so many question types that students are not sure how to use the computer to give an answer." (Jenny Hayes)

"Being able to review questions and change their responses." (Joan Gavin)

Twelve interviewees told us that students:

"Need to become familiar with the use of the technology during formative assessment." (Karl Wall)

Research confirms how important it is for students to develop e-assessment user skills:

"Almost all students feel that they are under-prepared when embarking on e-assessment tasks, be they formative or summative. It felt that because they use the internet and other communication technologies, such as mobile phones, socially, there is an assumption that they possess the necessary skills and do not require any substantive prior tuition." Learner perspectives of e-assessment, Project leader Matthew Badcock, School of Social Sciences, University of Central England.[7]

We also heard from interviewees about the characteristics of e-assessment that students dislike. A couple of people mentioned that students:

"Can be very stressed by an unfamiliar technology."

"Dislike e-assessment if they think academics are using it to reduce their student contact time."

Further, students:

"Hate it when the system crashes." (Jenny Hayes)

"Dislike scrolling screens." (Karl Wall)

[7]. http://www.c-sap.bham.ac.uk/ resources/project_reports/findings/ showfinding.htm?id=23/S/06

Students may show a lack of confidence in the computer system:

> "An unexpected reaction from students was to accuse the computer of deliberately marking their responses inaccurately." (Jenny Hayes)

However, this can be turned to the tutors' advantage:

> "In Further Education, lecturers have reported that poor results delivered through summative e-assessment led to students 'blaming the machine' (rather than themselves or the lecturers). This has facilitated a better relationship between lecturers and students in some cases, where the goal becomes working together to 'beat the machine'." (Kenji Lamb, JISC, N&E Scotland)

Lastly, we were told that students "like or hate getting results quickly depending on if they have passed or not." (Denise Whitelock)

For further information on the students' perspective, please see the National Union of Students' e-assessment principles provided in Appendix A.

Easy to use

As one interviewee told us, the feedback from their students is that high quality e-assessments have:

> "Clear, unambiguous questions – so that there is no need to work out what the question is asking nor any need to take guesses (provided that the candidate knows the subject/task well)." (Jenny Hayes)

Fair

With competency tests students place a high priority on fair e-assessments as failing the test can mean they lose their job or are denied a promotion. For these students fair e-assessments mean:

> "Questions that reflect the real job so they are asked in a way that is similar to the way that the candidate will deal with them on a day to day basis – not phrased esoterically." (Jenny Hayes)

These students consider the following characteristics contribute to an e-assessment being fair:

> "A clear view why s/he passed/failed the test e.g., a list of questions which s/he got wrong or partially wrong and which answer options s/he chose/entered.
>
> A test that has clear parameters, e.g., it states up-front how long the test is, any special passmark, any time limit, etc.
>
> A test that allows you to go back and review/change answers if you want to before finishing the test.
>
> Easy interaction (not too complicated dragging etc).
>
> Clear instructions (at both test and question level) so no doubts about what to do.
>
> A method of feeding back unhappiness about a result." (Jenny Hayes)

Addresses what has been taught

From a student's point of view it is important that:

"Questions are accurate and correct in terms of the subject and what has been taught." (Karl Wall)

Role: Someone who uses marks assessments

For competency tests, a student's manager wants:

"Confidence that if someone passes the test they have proved they have the essential underpinning knowledge for the job.

Clear advice about any areas of knowledge that require development and how important/urgent it is to ensure development occurs.

Easy access to the results of relevant tests – so can check candidate's portfolios and results.

Easy access to information about need for scheduling essential tests.

A way of feeding back if individual questions or whole tests do not seem to meet his/her needs (too hard to pass for example or one of the questions is out of date)." (Jenny Hayes)

1.2 Theoretical and international practice

Introduction

This section provides the received, expert point of view on "What denotes high quality in summative e-assessment". Following this survey of the state of the art, the next section critically compares practitioner practice from the previous section with this survey.

The section title reflects the fact that much of the research on and practice in summative e-assessment originates and takes place in the USA, and is the major source of expert opinion. There are also strong measurement cultures in the UK, Holland and Australia amongst other countries. Though its examples deal with school subject matter from the 1960's, Bloom's (1971) *Handbook*[8] remains the unsurpassed source of exemplar assessment of the sort considered in this report, addressing issues such as the assessment of higher-order cognitive skills, using objective tests in a wide variety of subject areas, systematic sampling of the subject domain and the intended learning outcomes, structuring tests so as to fairly reflect the widest range of abilities, and so on.

- For the purposes of this report, expert opinion on what constitutes high quality covers the following three areas, being the sub-sections which follow:
- Psychometrics: reliability, validity.
- Pedagogy: mapping to intended learning outcomes (ILOs), Gagne's skills levels, Bloom's taxonomy.
- Practicality: security, accessibility.

Psychometrics

Developed in detail starting from the 1920's work on intelligence and the 1940's work on personality, psychometrics is concerned with establishing how accurately and correctly a test measures what it says it measures. These issues are the test's reliability and validity. A recent introduction to the field of psychometrics may be found in Rust & Golombok's (2007) *Modern Psychometrics*[9].

In other contexts, these quality issues might be characterised as "V & V" (verify and validate): "does the test measure the thing right", and "does the test measure the right thing". Such characterisation is typical in the software engineering approach to quality (and analysed in texts such as Boehm & Turner's (2003) *Balancing Agility and Discipline*[10]).

The quality with which a test measures what it says it measures consists of its reliability and its validity.

---

[8] Bloom, BS, Hastings, JT, & Madaus, GF (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. McGraw-Hill.
[9] Rust, J & Golombok, S (2007). *Modern Psychometrics.* Routledge.
[10] Boehm, B & Turner, R (2003). *Balancing Agility and Discipline.* Routledge.

Reliability

In establishing that a test "measures the thing right", psychometric theory identifies a number of statistical measures which address the issue of a test's consistency, coherence, robustness, and reliability. With reference to typical summative e-assessments, reliability statistics answer questions such as the following:

- For the test as a whole, does it give the same result when different versions are used?
- Are the same results obtained when the test is re-taken?
- Do the "better" students do better on the test?

Test statistics should be reported by any reputable delivery software, and for a single instance of a test should report the Cronbach Alpha or the Kuder-Richardson formula 21 of overall test consistency. For comparing two instances of a test, relevant reliability statistics are given by calculating coefficients of correlation between the results of giving the same test twice (test-retest reliability) or of giving different versions to the same class (alternative forms reliability).

For the individual question items in a test, reliability statistics answer questions such as following:

- Do they show an appropriate range of difficulty?
- Is the selection of incorrect item options appropriately distributed?
- Do the "better" students tend to answer correctly?

As before, item statistics should be reported by any reputable delivery software as point-biserial or tetrachoric correlation coefficients to check that better students tended to answer the item correctly; as difficulty, discrimination, or facility coefficients to check the range of discrimination; and as option response histograms to check that each option was selected sometimes.

Validity

In establishing that a test "measures the right thing", psychometric theory identifies a further set of measures which address the issue of a test's validity. With reference to typical summative e-assessments, these validity measures and statistics answer the following:

- Do the test items follow directly from the intended learning outcomes?
- Do the test results predict future performance on similar tests?
- Do the test results correlate with other measures of student knowledge and ability?

Checking that the test items follow from the intended learning outcomes (ILOs) is a form of content or face validity. Checking that the test predicts performance is predictive validity. Checking that the test results correlate with other measures is a form of construct validity.

An interesting result from classical psychometric test theory is that the reliability of a test provides the upper bound to a test's validity. A test cannot be valid without also being reliable. Reliability in itself is insufficient, since a test can be reliable but not valid.

Appendix B provides more details of evaluating item and test quality.

Pedagogy

Pedagogical issues in e-assessment quality focus upon the verification of a student's achievement of an intended learning outcome (ILO).

We construe an intended learning outcome to be more or less co-extensive with what is also known as an educational objective, as well as with most structured definitions of "competence" or "competency". We note that while ILOs are elaborations of one or more educational aims or goals, such aims or goals cannot substitute for an ILO, since we define an ILO as comprising two components: a statement of a capability, and a statement of the subject matter to which that capability applies. An ILO is in principle prefaced by a phrase such as, "By the end of the module the student will be able to …", but this phrase is usually omitted in specifying an ILO since it is understood as being implied.

From this definition, it follows that an ILO is a statement about the sort of learning expected in Higher Education whose achievement is capable of verification. There is thus an isomorphism between an assessment and the ILO it tests, and it follows that an assessment may be defined as the verification of a capability with respect to an identified subject matter domain.

In considering the student's intended "capability", we are particularly careful to admit only "objectively" verifiable behaviours. Following Gagné, Bloom, and others, we note there are a number of taxonomies of cognitive, psychomotor, and affective behaviours which are considered "objective". We note in particular that there are a number of pseudo-statements of objective student behaviour which commonly circulate in Higher Education, such as "understand X", "appreciate Y", and "know Z", and that such pseudo-ILOs need appropriate elaboration before they may become amenable to the form of quality analysis contemplated here.

The point of such pedantry is to ensure that e-assessment quality can be judged at all. Without knowing what students should be able to do with their knowledge, it is simply not possible to properly assess their "possession" of such knowledge and, *a fortiori*, not possible to make any supportable statement about the quality of such assessment.

Assessments of pseudo-ILOs such as "understand X" are necessarily thus pseudo-assessments in this context. Assessments of students' "knowledge" or "appreciation" of some subject matter "Z" or "Y" are pseudo-assessments in the sense that such assessments are not readily amenable to evidence-based quality management.

Appendix C provides more details of the structure and content of ILOs.

Practicality

The practical issues considered by the report cover the two areas of security and accessibility.

Security

Security considerations in summative e-assessment are commonly divided into three categories, namely:

- Security of test materials.
- Security of assessment data transferred over networks.
- Security of results.

Appendix D provides more details of security standards and recommendations.

Accessibility

Most students, once they have become familiar with online testing, prefer this medium. In the past, guidance about making sure computer-based or online assessments are accessible to disabled students often avoided the technological issues that could arise and concentrated on the inclusion of extra time for completion or the provision of alternative formats such as paper-based examinations. There are situations where these suggestions may not always prove to be helpful and it is vital to be aware of individual preferences which have been highlighted in the checklist in Appendix E.

It is extremely beneficial for students to be allowed to go over the types of questions that will be asked, to be able to practice navigating around the online assessment site, and to make use of assistive technologies (such as screen readers or specialist input devices) to ensure access is possible, for example keyboard navigation within multiple choice questions.

> "The learner should engage with subject matter not the software. It is vitally important that the system does not end up assessing the student's ability to press buttons in the right order." Thomas & Milligan (2003)[11].
>
> "We need to ensure that every step of the way we are letting students know what is ahead of them in as much time as possible, so potential issuer and barriers can be identified, dealt with, and removed or avoided." Ball (2006)[12].

Appendix E provides more details of accessibility recommendations.

[11] Thomas, R. C., & Milligan, C. D. (2003). *Online assessment of practical experiments.* Proceedings of the Seventh International Computer Assisted Assessment Conference (pp. 421–430). Leicestershire, United Kingdom, Loughborough University. http://www.jelsim.org/resources/thomasandmilligan_caa2003.pdf

[12] Ball, S. (2006) *Accessible e-Assessment* CETIS-TechDis Accessibility SIG meeting http://zope.cetis.ac.uk/members/accessibility/meetings/2006/sig14/simonhtml#aims

1.3 Comparison between practitioner response and theoretical and international practice

Expert opinion on what constitutes high quality summative e-assessment focused, in this order of priority, on:

- Psychometrics (reliability, validity),
- Pedagogy (mapping to intended learning outcomes), and
- Practical issues (security, accessibility).

Interviewee comments on what constitutes high quality summative e-assessment fell, in this order of volume, into the categories of:

- Using the medium to give an extra dimension to assessment, including creating appropriate e-assessments;
- Issues around delivery including security, infrastructure reliability, and accessibility;
- Fairness;
- Ease of use;
- Supporting academic, managerial, and organisational goals;
- Addressing the intended learning outcomes; and
- Validity and reliability, mainly in their 'ordinary' senses, and sometimes in their psychometric senses.

In comparing these two lists, there are three outcomes that emerge.

The first is that the interviewees' list is longer and mentions issues not on the experts' list, such as 'using the medium to give an extra dimension to assessment', 'fairness', and 'ease of use'. We thus need to explain the relative brevity of the experts' list.

While interviewees mentioned 'fairness' and 'ease of use' as separate characteristics, the experts included these points in the characteristics of 'validity and reliability'. An invalid or unreliable test is unfair and an e-assessment system that is difficult to use would be invalid.

In the initial stage of the REAQ project, the expert panel considered a lengthy list of possible issues around the quality of summative e-assessment. On that list, for example, was a concern about an over-use of objective testing in certain courses in HE, and another concern that multiple choice objective tests may lead to impoverishment of assessment.

Comprehensively addressing these and the other questions was thought to require an investigation of the whole assessment regime, both conventional and e-, for a number of courses, looking in considerable detail at how tests and question items are in fact generated and come in practice to be incorporated into assessments.

After further discussion, the decision was taken to scope this report quite carefully, and to provide prioritised focus upon the essentials of quality in summative e-assessment.

The second outcome from the comparison is that the interviewees' list and the experts' list show complete overlap on the issues involved in high quality summative e-assessment, in the sense that every item on the experts' list appears on the interviewees' list.

This provides a pleasing congruence of expert and interviewee opinion, in that there is agreement on what the components of quality in summative e-assessment are.

The third, and most interesting, outcome is that the experts' views and the interviewees' comments show inverse priorities. While the experts saw psychometric measures as the first and most important feature in evaluating the quality of summative e-assessment, the interviewees did not accord this issue such significance, except for those who are learning technologists. Instead, we were surprised to see how delivery issues, and innovation issues, were so much more emphasised by the interviewees.

Quality as effectiveness

What our interviewees told us, at length, about e-assessment giving an extra dimension to assessment was valuable and interesting. We construe most of these comments as instantiating what our interviewees, their institutions, and funding bodies would call 'innovative' e-assessment.

While we are certainly enthusiastic about the potential of e-assessment, and are concerned that e-assessments are appropriate, we do not hold the view that 'good e-assessment is innovative e-assessment'. We would prefer to say simply that 'good e-assessment is effective e-assessment'.

Quality as psychometrics

In our view high quality e-assessment items need to be reliable and valid, yet few interviewees appear to be using psychometric measures of these factors.

It may be worth noting that there are some misconceptions about the appropriate role of item statistics, in particular the role of item difficulty (item facility) coefficients. A number of interviewees thought that that items should be rejected or selected based upon these values. This is a simplistic approach which is only partially true, and in any case may result in the rejection of excellent questions or the inclusion of completely inadequate questions.

Pedagogical quality

Addressing the intended learning outcomes was only discussed by a minority of people being interviewed, while many mentioned it is the responsibility of the academic involved. We could speculate on the reasons for this, but from the point of view of this report, the outcome is relatively clear: not many interviewees articulated the relationship between an assessment and the intended learning outcomes it assesses as a necessary component of quality.

Quality as reliable delivery

We were surprised that effective delivery was such an issue for quality in the interviewees' comments. We view delivery issues in e-assessment quality as 'hygiene factors', in that they may cause unsatisfactory assessments but cannot ensure quality in and of themselves. However, it seems that because e-assessment is not yet embedded in the exam structure and processes for many departments or institutions, delivery issues are a time-consuming concern for practitioners.

Other issues

Ease of use for the student (unambiguous questions and clear user interface) came out very strongly as a concern in producing quality e-assessment. We surmise that quality drops with poor tools and infrequent use of tools. This may be because good quality authoring tools are expensive and so is the time needed to allow academics/authors to become familiar with using them/making best use of them.

1.4 Tips and traps

The experts believe that the following points need to be addressed to create high quality e-assessment.

Tips

- Use psychometrics to manage reliability and validity.
- "Clearly mark every question with a unique identifier so that if a candidate or tutor/manager wants to query the validity of a question they can do so specifically." (Paul Booth)
- Place a high priority on security.
- Ensure that e-assessments are fair by making them accessible to all students, by ensuring the tests address the stated intended learning outcomes and that each question or part of a question is clear and unambiguous.
- Make good use of the medium by selecting testing methods and support material that are authentic to the skills being assessed.
- Consider using e-assessment to add another dimension to the summative test, for example by providing stepped questions, hints or collaborative tests. Authors need to ensure fairness when applying these techniques, for example by assigning partial credits. Other techniques that use the opportunities presented by 'e' are using multimedia, colour, drag and drop questions and matching questions. All of these are difficult to do on paper and offer benefits for summative assessment such as greater authenticity of the skills being tested and making tests easier to use.
- Have a good backup plan in place in case the technology fails, e.g., printed alternative papers, extra PCs available.
- Consider summative e-assessment as another opportunity for students to learn by providing tailored feedback. This is much easier to achieve with e-assessment than with traditional assessment methods as the feedback can be programmed into the e-assessment and provided to the students immediately or shortly after they sit the e-assessment.

  "Use the facilities of e-assessment to provide useful feedback to candidates about performance as individuals. Summative assessment needs to justify its 'scores' too. People are a lot less accepting now of the results handed down to them - they are more likely to accept an assessment if they can see where they went wrong (people don't usually query a good score!)" (Paul Booth)

Traps

- Although effective delivery will not improve the quality of your e-assessment's reliability or validity, poor delivery can make e-assessments unreliable and difficult to answer and mark. Processes that are specific to e-assessment must be developed such as ensuring questions will be displayed clearly on the monitors that the students will use and staggering the start times of exams so there are no problems with server delivery.
- Don't forget to apply the summative assessment good practice that has been developed for other methods to e-assessment, such as using appropriate language for the student and providing clear instructions. They still apply.
- "Beware of assuming that just because lots of people answer a question wrongly or score badly on a test, then there must be something wrong with the question or test. This is just one of the possibilities! (For example, one of the others is that the candidates were badly taught)." (Paul Booth)
- Piloting and checking e-assessments before they go live is vital. Common traps are leaving out these steps or not having enough time to make the changes such checks show are needed.
- Be careful about combining summative e-assessment with the open capabilities of a VLE. It may be better to separate them to avoid cheating.

2. Processes that produce high quality e-assessment

Interviewees were asked. 'What steps do you follow to create and use summative e-assessment?' Section 2.1 of this part focuses on what interviewees told us. Section 2.2 outlines the theoretical and international practice and guidelines for processes that create and sustain high quality e-assessment. Then section 3.3 compares what practitioners told us with what expert theory and practice suggests. The final section in this part, 3.4, pulls all the threads together by summarising the tips and traps in the quality management of summative e-assessment.

Our expectations are illustrated in Figure 9. We anticipated we would hear about quality management practice in terms of metrics, standards, process improvement, and capability maturity.

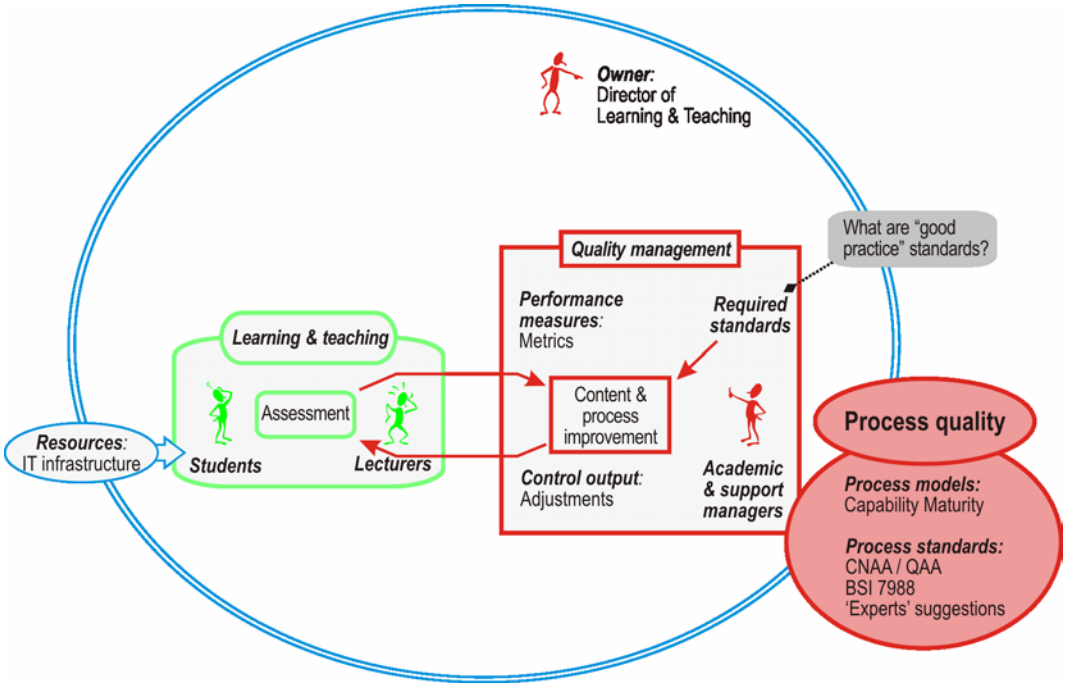

Figure 9: Project expectations of Part 2 issues

2.1 What interviewees told us

Interviewees were asked about the steps they followed to create and use summative e-assessment, about ensuring reliability, validity, security and accessibility, and about the differences with the processes involved in other types of assessment.

Comparison with traditional assessment methods

When asked how the process of creating high quality e-assessment differs from the process for creating high quality assessments using traditional methods, one of the interviewees said there were no differences and eight said there were no differences in principle, however,

"In practice there are major differences." (Ben Palmer)

And ten interviewees said that there were differences.

Assessment design

Many of these differences arise because of the different assessment designs that are possible in 'e'.

"The guidelines for creating good questions are the same [as traditional methods] but further types of questions are possible in e-assessment which offer chances and problems. There are chances to match the question type to the subject matter being assessed and there are extra possibilities of scoring to consider. Then there are logistical problems to solve such as creating good images, training invigilators and anticipating and solving problems." (Silvester Draaijer)

"It is difficult to get to grips with the potential offered by the software such as weighting questions and negative marking. Academics need to be aware of what their organization's tool can do. However, they need to beware of using aspects of the software they are not familiar with, such as negative marking." (Chris Ricketts)

"Reviewers need to be aware of what the software can do." (Helena Knowles)

Development processes needed

The other major difference is in the creation, checking, delivery and other processes needed to use summative e-assessment effectively:

"There are pitfalls and advantages that make e-assessment different from other methods, so you need different processes to address them. The process of creating and using e-assessment has to be designed from beginning to end." (John Kleeman)

"Long ago people thought through all the issues to do with paper-based summative assessments. We have good practice procedures for these types of tests. Now all that thinking needs to be done again for e-assessments". (Helen Ashton)

"The process of designing the software and testing of e-assessment validity and reliability add different aspects." (Robert Harding)

"E-assessment has flexibility which is good. But it means that the inputter has to understand how to set parameters to create variations of a question and then someone has to work through all the possibilities to check them." (Ben Palmer)

Several people mentioned that with traditional assessment if a question is badly worded the human marker and moderator can ensure that students are treated fairly. But with e-assessment you need to foresee all the problems up front.

Timing of the work needed

Interviewees also stressed that the timing of work needed to write e-assessments, their marking schemes and feedback is also up front, rather than after students have taken the exam. This adds to academics' workload at a time when they are already busy teaching. A key reason our interviewees cited for individuals and organisations

choosing e-assessment is because they expect it to save them time on marking. However,

> "This is a long-term benefit as the time needed to create the e-assessments in the first place is fairly high. The timing of when the questions need to be constructed is also an issue as this tends to be during busy times for academics." (Group discussion)

> "E-assessment creation and use is demanding on all parties. It takes time and requires new skills and new procedures. The role of Head of Department is critical to successful implementation."(Cliff Beevers)

> "Make sure that you have the resources to do the job well – it will take more effort than you think." (Ben Palmer)

Blurring the traditional boundaries between summative and formative assessments

> "I don't think there is any such thing as summative assessment. There is summative use of assessment data and formative use of assessment data (or both!)" (Helen Ashton)

> "The blurring of formative and summative is an essential theme as it means that the same questions can be used with different feedback applied to each to the benefit of all: efficient, helps learning, gives practice." (Cliff Beevers)

Steps taken in the creation and use of e-assessment

Interviewees responded with the following points when asked how they create high quality summative e-assessments.

Deciding how to use e-assessment

One interviewee explained how e-assessments could be designed:

> "Construct tests with students in mind, for example, test anxiety is common so start with easier questions and make later questions more difficult." (Karl Wall)

Others focused on the choice of question formats with one interviewee stressed that an e-assessment question must be an:

> "Excellent question in its own right, that works well in the medium." (Helena Knowles)

The learning technologists and more experienced academic advisers pointed out that:

> "Teaching staff are often unaware of the various options offered by e-assessment beyond the humble MCQ." (Kenji Lamb)

> "It is important to have someone with experience of the system to help write the questions. It is even more important to have the someone who knows the subject ... and who has taught." (Helen Ashton)

This point was reinforced by a learning technologist who said:

> "It is difficult for people who know psychometrics to work with subject matter experts to develop questions because each subject has complex issues. So we can only put forward general guidelines on how to design valid questions, such

as underlining the word not in a question and designing tests using a matrix of topics and levels of difficulty according to Bloom's taxonomy. It depends on the time and enthusiasm of the instructor whether they follow the guidelines or not." (Silvester Draaijer)

Team approach to the development of e-assessment

To address the complexities of each subject, interviewees recommended that managers and authors check assessments with subject matter experts and high performers (people who do the job well).

"A team approach is good for authoring questions, drawing on different perspectives and different ideas for implementation" (Nora Mogey)

"Question writing is a skill that needs to be learned. Write 10 questions each and then swap. People can learn a lot about good questions by trying to answer some and by having other people misunderstand what their questions meant to ask." (Paul Booth)

"You get more useful feedback from staff who are competent in the job and can see where students are likely to get confused." (Jenny Hayes)

"Round table discussion very useful." (Helena Knowles)

"A team approach may act as a catalyst to review existing questions" (Nora Mogey)

and checking them with people who are like the end users:

"Piloting is not done enough." (Karl Wall)

"Observe users and then ask them why they took the action they did." (Jenny Hayes)

"Ask users what they think the assumptions behind the questions are." (Karl Wall)

You can pilot questions as non-scoring components of an existing test. However:

"Be aware that students may spot this device and they tend to hate being used as guinea pigs." (Group discussion)

You can ask students who sit the assessments for their reactions:

"Students are often reluctant to make comments. Make an evaluation of the e-assessment a compulsory part of the course that is completed before results are issued." (Karl Wall)

"Note that students will always do things that you have not anticipated." (Ben Palmer)

External examiners who are not familiar with e-assessment will need support:

"With randomly-generated tests, externals cannot review all the 'actual papers' that students sat. They need help to understand how the computer has been instructed to devise tests from the pool of questions." (Group discussion)

Delivery

When it came to delivery, interviewees mentioned the following practical advice:

"Stagger the start of online exams – need to make sure invigilators know about this – to avoid overloading server." (Nora Mogey)

"Sometimes people are told the server can easily cope with 100 or 200 logons at the same time, unaware that the link between their students and the server is not as good. So, it's not just the server, it is also the connectivity that needs to be considered." (Paul Booth)

"Have paper copies of the exam as a contingency – about 10% of the number of students." (Jackie Lupton)

Gathering and using data from e-assessments

When it comes to deciding what you want to be able to do with the test results from your e-assessment tool, (such as review students, or classes, moderate questions, investigate common misconceptions etc):

"You can gather a lot of data that is never used because it in not in a usable format." (Group discussion)

"Piloting and 'iterate and refine' is important, but it's crucial to have a good reporting system in order to see and analyse the results (including student answers) in order to complete this process." (Helen Ashton)

Continuous improvement

"Continuous improvement is an essential part of ensuring high quality e-assessment. It is not always possible to predict that a test will be too difficult or a particular question will be ambiguous, because the people setting the test do so within a different context than the people who take the test. Also tests will need updating to reflect changes in real world." (Jenny Hayes)

"It is important to include feedback loops in all the steps of creating and maintaining e-assessments in order to sustain the process of continuous improvement. Feedback has to be encouraged and facilitated from all stakeholders in the e-assessment process and this includes candidates as well as authors and managers." (Paul Booth)

Content managers need QA tools

In particular we were told they require:

"An easy way of:

- analyzing test results both for individuals and for groups so as to identify 'weak' questions and improve/eliminate them and identify tests that are too hard or too easy and look at the reasons why in order to improve them.
- reviewing content to ensure that all aspects of the core knowledge for a competency can be seen to be covered by the assessment, i.e., ease of validation.
- identifying mandatory questions so that it is quick to check which questions are regarded as essential.
- changing questions and adding new ones/remove old ones quickly.
- change passmarks and results text to reflect changes in competence requirements or changes in development practices." (Jenny Hayes)

USA practice

Psychometric practice in USA

"I would say essential things that should/do get done are:

- Having sound item development processes where people that are creating the questions have sufficient knowledge in the domain being tested and items are peer reviewed before they are tested.

- Beta testing questions would be nice but realistically this is not always possible/practical in a post-secondary context.

- Questions should be thoroughly reviewed after they are initially tested to ensure that they do not have obviously psychometric flaws (e.g., negative CTT discrimination statistics). If questions are found to be flawed they should be dropped from the test and participants should have adjusted scores produced.

- Test level psychometric information should be evaluated especially the internal consistency reliability (e.g., Cronbach's Alpha).

- Cheating and item theft should be monitored as participants having prior knowledge of a question throws into question the validity of the results - cheater prevention methods should be used (e.g., shuffling question responses).

- Adding new questions and updating the item bank regularly to ensure items do not get stale and over exposed." (Greg Pope)

How people tend to start using e-assessment

Eleven of the interviewees told us that institutions tend to start by converting existing paper-based MCQs and other questions into 'e'.

"It is not enough to type the questions into your e-assessment tool. They need to be translated for the new medium. For example, making question stems simple and laying out the question clearly." (Group discussion)

"When converting a paper exam to an e-assessment format, questions typically have to be re-worded or changed in structure." (Martin Youngson, Heriot-Watt University)

"It can be painful. The unforgiving nature of 'e' means that authors need to review their assumptions about what makes a 'good' question and learn how to communicate in a new medium." (Group discussion)

"Training authors to use different question formats, as appropriate, can be difficult as they tend to resist examples [of good practice] that are not in their subject and at the appropriate level." (Silvester Draaijer)

Interviewees also had the following advice for anyone starting to use summative e-assessment:

"Start with formative e-assessment – it is safer for you and the students and can be reused." (Cliff Beevers)

"Don't be over ambitious, keep it as simple as you can and don't overload the students (do they really need to answer 50+ questions?)." (Ben Palmer)

> "Recommend email support in the initial stages of e-assessment implementation. It helps to smooth acceptance." (Bill Foster)

What can go wrong with e-assessment processes

Nine interviewees told us that the support allocated to academic staff for e-assessment tasks is often insufficient.

> "Colleges have set targets to digitise learning content or deliver e-assessment, but staff are often not given additional time (or sufficient training) to implement this." (Kenji Lamb)

Another result of limited time is that QA steps may be skipped. Also, staff may not have access to enough resources to deliver the e-assessment effectively.

Two interviewees told us that the time students spent on an e-assessment task during a timed exam was a problem. The assessment task involved creating an essay online. Students spent too long answering one question because they could spell-check their work, search the web etc. and so did not complete all the questions on the paper.

Twelve interviewees mentioned that:

> "Poor questions in a different medium will be poor questions in e-assessment. For example, unintentional clues in question stems and authors posing arcane questions that are not focused on what the assessment needs to achieve." (Jenny Hayes)

> "If you have got a rotten test as defined by quality of test items, no amount of clever trickery on the computer can make it a good assessment." (Robert Harding)

Four interviewees also highlighted that there is often:

> "Not enough trialling of questions on screens / system that will be used to deliver the e-assessment." (Bill Warburton)

Changes to the e-assessment technology was highlighted as a major problem by one interviewee:

> "The withdrawal of a version of an assessment engine leading to unusability of a colleague's many, many questions. This means we can be subject to the vagaries of the suppliers." (S7)

The technology also presented a challenge to another interviewee:

> "Exam supervisors are often people with limited technical skills. We need to train them how students should record their answers to e-assessment questions and how to move between questions. And at present our administrations systems have not caught up with e-assessment." (Jackie Lupton)

Ensuring reliability and validity

When asked 'How do you ensure reliability and validity?' we had a mixture of responses. Two interviewees said that reliability was the responsibility of their technical experts. Four others said that someone else was responsible for managing reliability and validity. Ten interviewees said that some academics have very limited knowledge of psychometrics. Their views were summed up in this comment:

> "Although many academics use the analysis reports very carefully, some are not fully aware of them or their importance and do not realise how or when to run them." (John Kleeman)

And:

> "Many instructors are only concerned with how many students passed or failed. We carry out the statistical analyses and advise the instructors on the action to take. Only when they know more about psychometrics are they interested in the reports [Assessment Analysis Report in v4.4 of Perception] themselves." (Silvester Draaijer)

Interviewees who specified psychometric measures used the following:

> "We measure how many people chose each option in a question. If more than 30% of candidates choose a distracter we investigate further. It may be an excellent distracter or the questions may be worded poorly." (Jenny Hayes)

> "We use facility values to determine level of difficulty – use FV to select questions for use in a test. The FVs can be refined through actual use." (Bobby Elliot)

> "Classical test theory and item analysis, reliability and quality measures such as correlation values, p-values and scores, measures of difficulty and cut score." (Silvester Draaijer)

> "A huge amount of analysis is done. We use descriptive statistics such as means, minimums and maximums. We measure the number of correct, incorrect and don't know answers per year. We analyse results using Cronbach's Alpha and item analysis. We analyse the test/retest reliability. We use two estimates of the standard error of measurement and test score distribution. We check for bias on test score for factors such as direct entry and mature students and locality." (Chris Ricketts)

> "The prize is to ensure a test can discriminate between people whose understanding is good/less good. The spread around a mean score is often a very good indicator." (Ben Palmer)

Eleven interviewees said that they expect the questions to be valid because the academic who teaches the course has set them.

However one interviewee recommended reviewing the questions against what has been taught (where possible) using peers and past students to make sure.

Ensuring security

Interviewees all consider security an important issue and mentioned the following as ways of ensuring it:

> "Secure your browser and be careful about using your usual LMS as it may not be secure enough." (Group discussion)

> "Limit access and number of attempts at an e-assessment." (Group discussion)

> "It is important to have your system show the student's identity at all times – this helps you check the student's identity when invigilating." (Helen Ashton)

> "Use screens as barriers to prevent students looking at each other's PCs." (Nora Mogey)

Allocating student desks and passwords is a common practice, as well as checking a student's identity against his or her university photo id card.

"It is possible to get systems that use webcams to take snapshots or video of the candidate taking an exam to be stored along with their results so that if there is any query about cheating by getting someone else to take the e-assessment, it can be resolved with photographic evidence. Additionally by publicising the use of webcams on the PCs used for test-taking it is possible to prevent cheating as people think that imposters are likely to be noticed and identified." (Paul Booth)

Ensuring accessibility

All interviewees mentioned using accessibility guidelines to ensure their e-assessments were fair to all students. Most of them also mentioned that they were not an expert themselves and relied on the literature for specific action to take.

"Follow guidelines and ensure that a step in the review process for questions, supporting documents/assets and tests looks at accessibility specifically." (Silvester Draaijer)

"Check timings needed for different groups." (Group Discussion)

"Check all equipment that is designed to aid students works in the test delivery environment." (Cliff Beevers)

"There are readability issues with different multimedia assets and methods of answering questions (such as drag and drop)." (Jenny Hayes)

2.2 Theoretical and international practice

Introduction

We first identify the steps that should be included in a complete end-to-end e-assessment cycle in UK HEIs.

Discussion around a variety of "capability maturity models" currently informs debates concerned with ensuring adequate processes for quality management and quality improvement. Originally developed by Watts Humphrey[13] at SRI to structure process improvements in software engineering, the concepts of capability maturity modelling have spread into a variety of other domains.

The tradition of proposing standards for products and processes has yielded documents which have some application to e-assessment quality management from the British Standards Institute (BSI -- BS 7988, now superceded by BS ISO/IEC 23988:2007), the International Test Commission (ITC), and the Quality Assurance Agency (QAA).

We also consider the current approach to Total Quality Management (TQM) for its relevance to this report.

Outline of steps

We believe that these points should be addressed when creating, delivering, and evaluating e-assessments:

1. Specify what the course assessment needs to achieve:

2. What expectations do you have of the student based on their learning and previous experience?

3. What is the learning outcome and purpose of the course? Or what is the job performance/competence to be assessed?

4. What is the context in which students need to demonstrate learning?

5. Specify the role that e-assessment will take in meeting this specification.

6. Breakdown the ILOs that the e-assessment will address into the appropriate underlying knowledge and skills.

7. Choose question formats and format of test.

8. Write questions and marking scheme. Set pass mark and write student feedback.

9. Check questions with subject matter experts / high performers.

10. Test questions work with your e-assessment delivery tool. For example, does the question work on the monitors you have? How long do questions / tests / audiovisual assets take to download and run?

11. Pilot questions with people who are like the end users.

12. Moderate the questions.

13. Make any changes needed.

14. Decide on the results feedback you want from your e-assessment tool.

[13] Humphrey, W. (1990). *Managing the Software Process*, Addison-Wesley.

15. Teach students strategies for answering e-assessments.
16. Run the e-assessment with real students and ask for comments.
17. Deliver the e-assessment.
18. Mark the assessments and/or compile the results of the exams.
19. Discuss with the external examiner and Exam Board.
20. Make any amendments required.
21. Publish the results.

Interviewees explained that practitioners need to follow all the steps in the process that are applicable to the e-assessment's circumstances. Problems can arise when steps (such as piloting the questions) are missed out.

Capability Maturity Models

The current variety of capability maturity models (CMMs) are rooted in the Carnegie-Mellon Software Engineering Institute's CMM Integration initiative[14]. The assessment maturity model[15] is a recent community development initiated by Question*mark* which identifies four levels of institutional capability.

Based upon the discussion in Gilbert & Gale[16], a possible CMM for summative e-assessment would suggest that institutional capability would be related to the use of:

- E-assessment management – knowing what the process tasks are, how long they take, what resources they require, and their appropriate sequencing and concurrency.
- Standards – adopting standards relevant to e-assessment such as reliability, validity, security, and pedagogic appropriateness.
- Metrics – measuring conformance to the standards adopted (reliability statistics such as Cronbach alpha, validity measures such as the proportion of verifiable links between ILOs and e-assessment items) and achievement of the e-assessment management tasks.
- Continuous process improvement – improving the e-assessment process maturity through refinement of the management tasks, development of the standards applied, and measuring and evaluating the effects of improvement initiatives.

Appendix F provides more details of such an e-assessment CMM.

Standards

British Standard ISO/IEC 23988:2007

The best known standard for delivering e-assessment in the UK is BS ISO/IEC 23988:2007 which recently superceded BS 7988 (2002). It makes recommendations for using IT to deliver assessments and then to record and score participants' responses. Its recommendations are in terms of assessment types, assessment life cycle stages and specific IT aspects.

[14] http://www.sei.cmu.edu/cmmi/general/index.html
[15] https://assessmentmaturitymodel.wikispaces.com/
[16] Gilbert, L, & Gale, V (2008). *Principles of e-Learning Systems Engineering*. Chandos.

BS7988 was originally developed by a panel of experts in the field of computerised assessment including representatives from government, business, education and exam boards and was published in 2002. It provides a set of guidelines on good practice for organisations that use or deliver computerised assessments and uses a commercial training model to define the responsibilities of stakeholders (assessment sponsors, assessment distributors and assessment centres). These commercial terms must be translated to an academic context (institution, academic support services, programme team).

Section 4 outlines some key guiding principles underlying the use of e-assessment, including the importance of ensuring that assessments are valid, reliable and fair to all candidates (including those with disabilities) and preserving the confidentiality and integrity of the assessment content and student data at all stages. Other principles are to maintain an audit trail (enabling queries to be investigated) and to make contingency plans such as back-up facilities and fall-back procedures to minimise the impact of technical failures and other problems.

BS ISO/IEC 23988:2007 also lists some guidelines covering the responsibilities of sponsors deciding to use CAA and technical/usability requirements of e-assessment systems. In addition it provides guidance on provision of scores and feedback to participants, secure management of questions and results, provision of information and practice tests to participants, the specification of assessment centres, preparation for and conduct of e-assessments and procedural advice on managing emergencies, technical failures and other irregularities.

ITC's (2005) Guidelines on Computer-Based and Internet Delivered Testing

The International Test Commission (ITC) published its Computer-based and Internet delivered testing Guidelines in 2005: they are comprehensive, if focused somewhat on commercial assessment. There is much in them to inform the quality assurance of e-assessment. They are available online[17] and provide detailed and authoritative guidance for Test Developers, Test Publishers and Test Users.

QAA (1994) Guidelines on quality assurance

The Further and Higher Education Act 1992 (F&HE Act 1992) caused the CNAA to be wound up and the CVCP to be expanded. It also saw the emergence of the QAA and made a single set of QA guidelines appropriate.

Appendix G provides more details of the QAA (1994) guidelines and Appendix H summarises the guidelines for e-assessment.

QAA (2006) Code of practice for the assurance of academic quality and standards in higher education

The original QAA Code of Practice's 'precepts and guidance' format was criticized by Boyle and O'Hare (2003) for giving what could be taken as vague statements of intent rather than prescriptions for implementation by HEIs. The QAA went some way to address this criticism in 2006 with the publication of a revised Code of Practice which explicitly adopts a 'precepts and explanation' approach, using explanations to clarify why the precepts are considered important (and thus reducing, as the authors

[17] February 2009, at http://www.intestcom.org/_guidelines/guidelines/index.html

say, opportunities for a 'checklist' approach). Practical prescriptive advice is not the aim of the revised QAA Code of Practice. BS ISO/IEC 23988:2007 combined with IEC's Guidelines on Computer-Based and Internet Delivered Testing is perhaps the best source of that.

Total quality management

Total quality management (TQM) is the term used to describe the process and management of change, designed to realign institutional mission, culture, and working practices to the pursuit of continued improved quality. A TQM culture is predicated upon a commitment to customers' needs, interests, requirements and expectations, and upon the commitment of everyone to the constant improvement of everything that an organisation does and provides for its customers. TQM has five guiding principles:

- Creation of an appropriate institutional climate;
- Focus placed on the (internal as well as external) customers;
- Evidence-based management;
- People-based management; and
- Continuous quality improvement.

In introducing TQM to an institution, emphasis is put on training senior managers first because they initiate the changes required. Training is then cascaded to the rest of the institution, concentrating first on staff with any co-ordinating, leading, or management role.

There is an extensive literature on the introduction, application, sustainability, and evaluation of TQM in Higher Education. From the point of view of this report, the value and virtue of TQM is that it provides the leading articulated expression of how a UK HE institution might think about ensuring e-assessment quality.

Trajectories/dual path

Patterns of e-assessment use can be considered as a special case of assessment lifecycles. They can be presented as trajectories categorized according to tutors' previous experience and risk propensity (Figure 10), particularly where credit-bearing assessment is concerned (Warburton 2009).

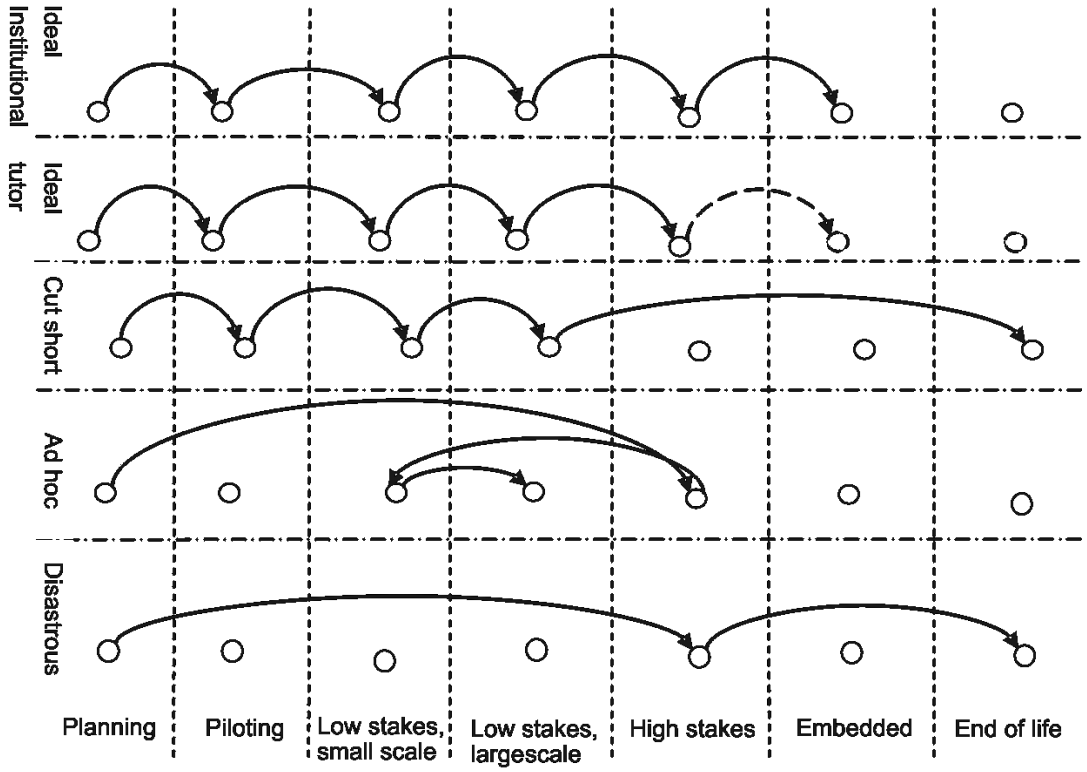

Figure 10 - E-assessment uptake trajectories

Warburton found that the biggest influences on tutor trajectories to be their motives for using e-assessment: where the aim was primarily to improve learning and teaching practice, the consequence was more likely to be a sustained progression through the different stages of e-assessment use that resulted in lower risk trajectories. Conversely, tutors who concentrated on productivity benefits and aimed directly for credit-bearing summative assessment take substantially greater risks and the probability of a non-optimal outcome was much increased. This so-called 'dual path' model of uptake helps to explain reports of mixed results when rolling out e-assessment in universities (Appendix I).

Reference

Warburton, W.I. (2009). Quick Win or Slow Burn? Modeling UK HE CAA uptake. In *Assessment and Evaluation in Higher Education,* 34(03).

2.3 Comparison between practitioner response and theoretical and international practice

Expert opinion on what constitutes quality management processes for summative e-assessment focused on:

- Establishing sets of steps to follow for both content and quality management.
- Identifying, using, and developing relevant standards for both content and quality management.
- Identifying metrics for both content and process.
- Capability maturity modelling as an encapsulation of these three essential elements of a quality management process.

Interviewee comments on what constitutes quality management processes for summative e-assessment fell under a variety of rules of thumb or suggestions for useful steps, such as: noting that the effort needed to write e-assessments, their marking schemes, and to construct feedback is front-loaded; starting with easier questions and making later questions more difficult; checking assessments with subject matter experts and high performers; identifying 'weak' questions and improving or eliminating them; reviewing content to ensure coverage; getting help for academics who usually have very limited knowledge of psychometrics; attending to security; using accessibility guidelines. In summary, these comprised:

- Heuristic steps for both content and quality management.
- Accessibility standards.

In comparing these two lists, there are two outcomes that emerge.

The first is that the experts' list is longer and mentions issues not on the interviewees' list, such as 'using standards for both content and process', 'metrics', and 'capability maturity modelling'.

Related to this is the second outcome: that the interviewees' list and the experts' list show only partial overlap on the issues involved in the quality management of summative e-assessment.

Institutions may apply the 'standard' processes used to assure the quality of traditional assessment to e-assessment, (such as checking exams internally and externally with experts before use), and therefore interviewees did not mention them specifically when asked. However, we wonder if summative e-assessment practitioners, in coming to reclaim the swamp, are often engaged in fighting alligators, and may not have clear systems for or systemization of quality management for e-assessment.

Capability maturity modelling

A capability maturity model (CMM) is likely to emphasise at least the three aspects of quality management identified in the experts' list, being the importance of having, for both content and quality management, a systematized process, deployed standards, and metrics. While we did not expect many interviewees to articulate a CMM for their quality management, we did expect more awareness and use of the essential ingredients of a CMM. In particular, we expected to hear more about the use of standards, either external or internally developed and tuned, and we hoped to hear of the use of metrics for quality management.

2.4 Tips and traps

Tips

- Use a team approach to the development of e-assessments to ensure that subject-matter expertise, e-assessment expertise and IT delivery expertise is included.
- Include feedback loops in all the steps of creating, using and maintaining e-assessment to enable continuous improvement.
- Pilot test e-assessments with students who are similar to the candidates who will sit the e-assessment.
- Help external examiners understand how the e-assessment works and what the different questions that will be generated will contain.
- “Require a 'verifier' to unlock high status tests so that if a candidate wants to cheat, they need to 'conspire' with the verifier (invigilator) as this reduces the likelihood of occurrence. Put the verifier's name next to the result as well as the candidate so that it is on public record as this also creates an obligation to be honest.” (Paul Booth)

Traps

- “It is possible to become too concerned with 'security' of e-assessments and it is important to keep the level/importance of cheating in perspective.” (Paul Booth)
- There is a danger of becoming too wrapped up in the best practice, which stifles the ability to get service up and running. Conversely, some institutions may have adopted an “alligator fighting” approach that has served them well up to now, but may well be hampering expansion of services.

3. Recommendations

The key differences between summative assessment using traditional methods and e-assessment have emerged from this survey. While traditional methods are well understood and the processes needed to produce high quality are known, it is still early days in the understanding of what high quality e-assessment is and what processes will ensure it is produced.

We learned that there is no generally accepted definition of what is meant by high quality summative e-assessment. However there are many characteristics that interviewees mentioned in common such as using the medium to add an extra dimension, authentic, accessible and fair, secure and a good discriminator between more and less able students. We expected to hear about the use of psychometrics to ensure quality. Although some people are using these measures, many are not. Instead we heard more about the importance of infrastructure and practical arrangements for successful e-assessment implementation. This reinforces the view that 'e' specific guidelines are needed to replace the well-known approaches already in place for traditional methods. We also found that because it is easier to provide feedback and hints in 'e' than with traditional methods, the traditional boundaries between summative and formative assessments are becoming blurred. We heard about the benefits of a team approach to creating e-assessments. And we learned that although there are pockets of good practice for the creation and use of e-assessment within certain schools, departments, subjects or individuals, overall institutions or organizations are at early stages on the path of implementing 'e'.

We therefore make the following recommendations for future work which JISC or other members of the community may wish to sponsor:

- A toolkit for the end-to-end process of creating e-assessment should be developed.
- A practical guide to the steps involved in creating and maintaining an e-assessment system.
- Guidelines for the quality assurance of e-assessments.
- Psychometric measures for assessing the quality of item banks rather than individual questions, for assessing, tracking, and reporting the quality of banked items during their lifecycle of use.
- Development and extension of existing psychometric theory to include multi-staged and optional stepped constructed response questions.
- Workshops and support materials to disseminate good practice in the use of psychometrics for selected response items and for questions employing constructed responses.
- Workshops and support materials to disseminate good practice in question creation and meeting educational needs beyond simple selected response items, possibly subject based.
- Accessibility and user interface guidelines for deploying e-assessment, in particular addressing the use of browsers.
- Guidelines for the use and role of MathML for expression recognition in e-assessments.
- A repository of exemplars of good practice for both selected response and constructed response questions.

The following recommendations are for developments in current practice for JISC or the community:

- JISC and other community calls for and sponsorship of e-assessment bids should consider where and how bidders should incorporate appropriate psychometric measures in their proposals.
- Commercial vendors should improve the accessibility of their psychometric reports to all stakeholders, possibly simplifying them to encourage take-up of their contents.

Further reading

The following were recommended to us by project participants.

Question design

Ashton, H (2008). Exploring e-Assessment: A practitioner's perspective. Retrieved February 2009 from http://www.elearning.ac.uk/features/eassessment

The Scottish Qualification Authority publishes a 'Guide to Writing Objective Tests', contact Bobby Elliot

Draaijer, S., & Hartog, R. (2007). Design patterns for digital item types in higher education. *e-Journal of Instructional Science and Technology*, 10(1). http://www.usq.edu.au/electpub/e-jist/docs/vol10_no1/papers/full_papers/draaijer_hartog.htm

Draaijer, S., & Hartog, R. (2007). Guidelines for the design of digital closed questions for assessment and learning in higher education. *e-Journal of Instructional Science and Technology*, 10(1). http://www.usq.edu.au/electpub/e-jist/docs/vol10_no1/papers/full_papers/draaijer_hartog_hofstee.htm

Hartog, R., Draaijer, S., & Rietveld, L.C. Practical aspects of task allocation in design and development of digital closed questions in higher education. *Practical Assessment, Research and Evaluation* (PARE). http://pareonline.net/getvn.asp?v=13&n=2

Flexibilization of Assessment: https://www.surfgroepen.nl/sites/flexassessment/Shared%20Documents/Handbook%20Towards%20Flexibility.aspx

Cynthia Parshall – Design of Innovative Items: http://www.cbtconference.com/images/presentations/CParshall.pdf

Blue Print: http://www.caacentre.ac.uk/bp/

For distance learners:

Ross, S.M, Jordan, S.E and Butcher, P.G. (2006). Online instantaneous and targeted feedback for remote learners, in Bryan, C. and Clegg, K.V. (Eds), *Innovative assessment in Higher Education*, Routledge, ISBN 0415356415

Jordan, S.E.; Butcher, P.G.; Ross, S.M. (2003). Mathematics assessment at a distance, at http://ltsn.mathstore.ac.uk/articles/maths-caa-series/july2003/index.shtml

Examples of good questions

Ashton, HS & Youngson, MA (2004). Creating questions for automatic assessment in mathematics. Retrieved February 2009 from http://ltsn.mathstore.ac.uk/articles/maths-caa-series/feb2004/index.shtml

Fawcett, L, Foster, B, & Youd, A (2008. Using computer based assessments in a large statistics service course. *MSOR Connections*, 8(3), 45-48.

Foster, W H (2004). Keeping control using computer aided assessment. Retrieved February 2009 from http://www.mathstore.ac.uk/articles/maths-caa-series/dec2004/index.shtml

Foster, B (2007). Using computer based assessment in first year mathematics and statistics degree courses at Newcastle University. *MSOR Connections*, 7(3), 41-45

Ashton, H (2008). Exploring e-Assessment: A practitioner's perspective. Retrieved February 2009 from http://www.elearning.ac.uk/features/e-assessment

Korabinski AA, Youngson, MA, & McAlpine, M (2008). Estimation of the accessibility of items and the confidence of candidates: A Rasch-based approach. *J. App. Measurement*, 9(1).

Sample question at http://www.pms.ac.uk/pms/undergraduate/assessment.php

From the Open University, UK:

http://www.open.ac.uk/openmarkexamples/index.shtml

https://students.open.ac.uk/openmark/omdemo.iat/

Case studies of e-assessment

Whitelock D. (2005). E-Assessment Case Studies: Update. Retrieved February 2009 from http://www.jisc.ac.uk/media/documents/programmes/elearningpedagogy/eassessmtcasestudies.pdf

JISC, Effective Practice with e-Assessment, An overview of technologies, policies and practice in further and higher education. Retrieved February 2009 from: http://www.jisc.ac.uk/media/documents/themes/elearning/effpraceassess.pdf

Security

E. Shepherd, J. Kleeman, J. Phaup, K. Fair and M. Belton: Delivering Computerized Assessments Safely and Securely. A Question*mark* White Paper retrieved February 2009 from: http://www.questionmark.com/uk/whitepapers/index.aspx

Measuring Learning Results

Thalheimer, W. (2007, April). Measuring Learning Results: Creating fair and valid assessments by considering findings from fundamental learning research. A Question*mark* White Paper retrieved February 2009, from http://www.questionmark.com/uk/whitepapers/index.aspx

Defensible Assessments

G. Pope, Defensible Assessments: What You Need to Know. A Question*mark* White Paper retrieved February 2009, from http://www.questionmark.com/uk/whitepapers/index.aspx

Student perspective on e-assessment

There is the Learner Perspective on e-assessment project:

http://www.c-sap.bham.ac.uk/resources/project_reports/findings/showfinding.htm?id=23/S/06

There is earlier work on student perceptions of assessment which covers attitudes to MCQs:

Struyven, Katrien, Dochy, Filip & Janssens, Steven (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, 30 (4), 325-341

Impact on students:

Northcote, M (2003). Online assessment in higher education: The influence of pedagogy on the construction of students' epistemologies. *Issues In Educational Research*, Vol 13. Available at: http://www.iier.org.au/iier13/northcote.html

This looked at children but may be of interest:

Johnson, M. & Green, S. (2004), 'On-line assessment: the impact of mode on students' strategies, perceptions and behaviours'. Available at: http://www.cambridge-assessment.org.uk/ca/digitalAssets/113885_On-line_Assessment._The_Impact_of_Mode_on_Students__Strategi.pdf

(They also published a version of this as Johnson, M. & Green, S. (2006). On-Line Mathematics Assessment: The Impact of Mode on Performance and Question Answering Strategies. *Journal of Technology, Learning, and Assessment*, 4(5). Available from http://www.jtla.org)

This is on formative assessment, but may be of value:

Walker, D. J.; Topping, K. & Rodrigues, S. (2008). Student reflections on formative e-assessment: expectations and perceptions, *Learning, Media and Technology*, 33, 221 - 234.

References

Quality Assurance

American Educational Research Association (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Copies of BS9788 are available from the British Standards Institution: www.bsi-global.com

Boyle, A. & O'Hare, D. (2003). Finding Appropriate Methods to Assure Quality Computer-Base Assessment in UK Higher Education. Online document found (January 2009) at http://www.caaconference.com/pastConferences/2003/procedings/boyle.pdf

ITC (2005). Computer-based and Internet delivered testing Guidelines.

Moore, G. A. (1999). *Crossing the chasm* (2nd ed). Oxford: Capstone.

QAA (2006). Code of practice for the assurance of academic quality and standards in higher education Section 6: Assessment of students - September 2006. Online document found (January 2009) at http://www.qaa.ac.uk/academicinfrastructure/codeOfPractice/section6/COP_AOS.pdf

Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed). New York: Free Press.

Shephard K., Warburton B., Maier P. & Warren A. (2006). Development and evaluation of computer-assisted assessment in higher education in relation to BS7988. *Assessment & Evaluation in Higher Education* Vol. 31, No. 5, October 2006, pp. 583–595. Link to article: http://dx.doi.org/10.1080/02602930600679621

Crawford, F (1991). A CVCP Occasional Paper; Total Quality Management. CVCP, London.

Sherr, L A & Teeter, D J (1991). Total Quality Management in Higher Education. *New Directions for Institutional Research* (71), Jossey Bass Inc.

ITC (2005). International Guidelines on Computer-Based and Internet Delivered Testing. Online document found (January 2009) at: http://www.ase-solutions.co.uk/downloads/ITC_Guidelines_Computer_based.pdf

Warburton, W. (2008). Quick Win or Slow Burn: Modeling UK HE CAA uptake, Assessment & Evaluation in Higher Education iFirst Article. Online document found (January 2009) at: http://dx.doi.org/10.1080/02602930802071080

Appendices

Appendix A: National Union of Students' Principles of Effective Assessment[18]

Principle	Comment
Should be for learning, not simply of learning	This positions assessment at the heart of learning rather than it serving as a simple add-on at the end of the process.
Should be reliable, valid, fair and consistent	It is crucial for staff, students and employers to have confidence in the assessment processes and their outcomes.
Should consist of effective and constructive feedback	Effective feedback on assessment is a crucial aspect of assessment processes and a key feature of enhancing the learning process.
Should be innovative and have the capacity to inspire and motivate	Formative assessment practices have the potential to inspire and motivate, and this aspect can be captured by innovative approaches, including those making use of new technology
Should measure understanding and application, rather than technique and memory	Assessments need to have a holistic approach that transcends the particular method being used; only this will truly test and reflect levels of learning.
Should be conducted throughout the course, rather than being positioned as a final event	Positioning assessment as an integral part of the course helps facilitate continuous learning.
Should develop key skills such as peer and reflective assessment	Not only do such mechanisms allow students to receive extra feedback on work beyond that of their tutor, they also help develop the key skill of self-reflection.
Should be central to staff development and teaching strategies, and frequently reviewed	Assessment processes must be innovative and responsive to learners' needs, and as such they need to be central to staff development and teaching strategies.
Should be of a manageable amount for both tutors and students	While assessment should be place in a central role in learning, for it to be effective neither tutor nor student should be overburdened.
Should encourage dialogue between students and their tutors and students and their peers	It is important that students and staff share the same definitions and ideas around standards. This can be fostered by increased dialogue and engagement.

[18] Times Higher Education, 29th January 2009

Appendix B: Evaluating item and test quality[19]

Difficulty level

The difficulty level of an item is simply the proportion of students taking the test who answered the item correctly. Items with a range of difficulty levels are needed in an MCQ test. While most items (perhaps 70 per cent) should show difficulties between 0.4 and 0.6, the rest should show difficulties as low as 0.85 and as high as 0.15. This is to enhance discrimination between students in the upper and lower quartiles. A 'good' item is not judged on its difficulty alone.

Distracter attractiveness

Each of the distracters should be roughly equally attractive, and a histogram or count of the number of students choosing each distracter should show a uniform distribution. For example, it would be a poor MCQ item if it had a distracter which failed to be the choice of any student.

Item-test correlation

The most important measure of a 'good' item is its correlation with the student's overall mark. That is, a student who did well on the assessment as a whole should tend to get the item right, and a student who did poorly overall should tend to get the item wrong. The correlation is usually measured using the point-biserial coefficient r (identical to the Pearson product-moment r when one of the variables is dichotomous). An analogue of the item-test correlation can also be measured by the inter-quartile range, or a t-test, between the mean mark of those who pass and those who fail the item.

Test psychometrics

There are established criteria for evaluating the quality of a test of knowledge, traits, skills or competencies, the most important being test reliability and test validity. These are summarised in the following table, Table 1.

Psychometric consideration	Characteristic	Measure
Reliability	Internal consistency	Correlation between items
	Alternative forms	Correlation between alternates
	Test-retest	Correlation between instances
	Inter-rater	Correlation between scorers
Validity	Content	Inspection of declared aims & actual content
	Predictive	Correlation with outcome
	Construct	Correlation with other measures of the same skill or competence

[19] Extract from Gilbert & Gale (2008), *Principles of E-Learning Systems Engineering.*

Table 1 Psychometric evaluation of tests

Reliability

Test reliability takes a number of forms. One of the most useful, and one of the easiest to measure, is the internal consistency of the test items. Popular statistics for measuring internal consistency include Cronbach's Alpha and Kuder-Richardson formula 20. One of the strengths of computer-scorable tests is that their inter-rater reliability measure is 1.0, i.e., perfect.

Validity

Test validity takes three forms, known as content or face validity, predictive validity and construct validity. Establishing the specific validity of a student test is rarely undertaken, perhaps because the test is not usually stable over time, and the cost of demonstrating or achieving adequate validity could well be equal to the cost of the entire e-learning unit or programme of which it is a part.

Appendix C: Intended learning outcomes[20]

Bloom's taxonomy of educational objectives

The analysis and design of objectives is associated with Bloom (1956) and Mager (1997). Bloom identified a taxonomy of educational objectives, illustrated below Figure 11, involving three domains: cognitive, affective, and psychomotor.

COGNITIVE DOMAIN	AFFECTIVE DOMAIN	PSYCHOMOTOR DOMAIN
Knowledge	Receiving	Imitation
Comprehension	Responding	Manipulation
Application	Valuing	Precision
Analysis	Organisation	Articulation
Synthesis	Value complex	Naturalization
Evaluation		

Figure 11: Bloom's domains

There are theoretical and practical problems with Bloom's hierarchy within the cognitive domain, and it is currently more fashionable to identify three broad areas of ability within this domain: "to be able to recall knowledge", "to be able to use knowledge", and "to be able to find new knowledge". The terminology of 'educational objectives' has also changed to 'intended learning outcomes' or ILOs.

Intended learning outcomes

The basic form of an ILO is specified below.

"The student will be able to X"
where X is a performance.

Figure 12: The form of a basic ILO

Whether explicitly stated or merely implied, an ILO begins with a standard phrase "By the end of the course, the student will be able to...". A statement of the performance that the student will be able to undertake completes the ILO. The performance involves an ability that the student has learned, expressed in behavioural terms. This means it must be possible to observe and assess whether the student can actually exhibit the ability in question. It is exactly this feature of an ILO which makes it central to assessment and e-assessment in particular.

[20] Extract from Gilbert & Gale (2008), *Principles of E-Learning Systems Engineering.*

The statement of the performance begins with the ability and concludes with one or more measurable or assessable actions by which it can be observed that the student has learned the ability in question. The ability generally comprises a verb denoting the learned capability followed by the object or subject matter of the learned capability. The table below, Table 2 lists these basic components of an ILO along with an example of each.

Component	Example
ILO	The student will be able to analyse target audience characteristics by listing those characteristics pertinent to the e-learning under consideration.
Performance	… to analyse target audience characteristics by listing ….
Ability	…analyse target audience characteristics…
Learned capability verb (LCV)	…analyse…
Object (subject matter)	…target audience characteristics…
Assessable behaviour	…listing those characteristics pertinent to the e-learning under consideration.

Table 2: Components of a basic ILO

The statement of a performance in terms of an ability and assessable behaviour with respect to some subject matter content provides the minimum form of an ILO. It may be important, however, to more closely define what is intended as the learning outcome, and attention may then be given to the supplementary components of a fully specified ILO: the situation or scenario within which the student needs to be able to perform; the standards of the performance; any constraints which may apply to the performance; and the use of any tools.

Capability	Standard LCV	Assessable behaviours
Knows facts	Recalls	Writing, drawing, indicating
Knows concepts	Defines	Writing, selecting
Knows procedures	States steps	Writing, drawing, flowcharting
Knows principles	States cause-effect relationships	Writing, drawing, graphing, specifying formula
Uses concepts	Classifies	Writing, selecting, sorting, arranging
Uses procedures	Demonstrates	Manipulating, calculating, measuring, constructing
Uses principles	Predicts	Calculating, drawing, graphing
Finds concepts	Invents	Sorting, specifying
Finds procedures	Devises	Experimenting, analysing
Finds principles	Discovers	Experimenting, analysing, observing, demonstrating

Table 3: Observable ability verbs after Merrill

Reference to Bloom's work identifies a rich set of verbs that can be used as required for both the learned capability and for the assessable behaviours by which that ability is demonstrated. These are shown in the following tables, for the cognitive domain, affective and psychomotor domains.

	Cognitive ability	Assessable behaviour
Knowledge	Knows terms, specific facts, rules, trends, categories, criteria, methods, procedures, principles, concepts, theories.	Name, label, define, state, recognise, list, recall, identify
Comprehension	Translates and paraphrases communications; interprets, summarises, and explains relationships; extrapolates from given data.	Explain, classify, summarise, extrapolate, interpret, convert
Application	Applies concepts, principles, rules, procedures.	Calculate, solve, construct, use, prepare, predict, demonstrate
Analysis	Analyses elements, relationships, or organisational principles; analyses connections, relationships, or arrangements.	Compare, contrast, infer, explain
Synthesis	Produces new arrangement or new result.	Compose, originate, design, create
Evaluation	Judges on the basis of criteria and evidence.	Appraise, argue, evaluate, criticise, assess, discriminate

Table 4: Performance verbs for the cognitive domain after Bloom

	Affective state	Assessable behaviour
Receiving	Indicates attention, shows willingness to receive, awareness of distinctive features	Describes, identifies, selects, demonstrates
Responding	Accepts need for regulations & responsibilities, chooses to respond in acceptable ways, shows satisfaction in responding	Demonstrates, tells, performs
Valuing	Endorses propositions, shows preference for values, expresses commitment to values	Explains, justifies, defends, demonstrates
Organisation	Understands relationship between values, develops a value system	Explains, defends, justifies
Value complex	Acts in ways consistent with values	Displays, practices, demonstrates

Table 5: Performance verbs for the affective domain after Bloom & Krathwohl

	Psychomotor state	Assessable behaviour
Imitation	Observe a skill and attempt to repeat it	Imitate, copy, replicate, duplicate, begin, attempt, follow, mimic
Manipulation	Reproduce a skill or movement sequence from instruction; Includes trial and error until an appropriate response is achieved	Perform, execute, build, re-create, implement, manipulate, improve
Precision	Perform a skill or movement sequence independently and proficiently with an emphasis on accuracy, proportion and exactness	Complete, achieve, demonstrate, show, control
Articulation	Modify the skill or produce the product to fit new situations; Combine more than one skill in sequence with harmony and consistency	Adapt, alter, combine, integrate, develop, re-arrange
Naturalization	Use skills or movement patterns that are now autonomous; Begins to experiment or create new motor acts	Perform automatically, invent

Table 6: Performance verbs for the psychomotor domain after Dave

Level of skill difficulty - Gagné's hierarchy

The following table, Table 7 shows the hierarchy of intellectual skills developed by Gagné with the simplest skill at the bottom. The most complex intellectual skill is cognitive strategy that combines all the skills into a coherent approach. The simplest skill is to be able to recognise an item when it is presented on its own.

Level of difficulty	Example
Cognitive strategy	Application of strategy involving rules, concepts etc.
Higher-Order relationship rules	Generation of a higher order rule through problem-solving, decision making or application of the rule to a problem
Relationship rules	Use of relationship between concepts so student can demonstrate rule
Defined concept	Definition of relationship between items or events. Classification of given items. Application of concept
Concrete concept	Identification of class of item
Discrimination	Identification of item from list of possibles. Specifying sameness and differences between examples
Recall and repeat	Recitation of definitions, set phrases etc. from memory
Recognise	Recognition of item when presented with it alone

Table 7: Gagné's hierarchy of skills

Appendix D: Security

The following recommendations are based on BS ISO/IEC 23988:2007 and the ITC's (2005) Guidelines on Computer-Based and Internet Delivered Testing which are both widely accepted as delineating good practice within the e-assessment community.

(a) Security of test materials

Responsibilities of Authors

Authors must minimise the risk of test items, scoring keys, and interpretation algorithms being illegitimately printed, downloaded, copied, or sent to another computer. In HE environments, this is principally facilitated by tutors being security conscious and liaising with IT departments to make good use of secure IT infrastructures.

Responsibilities of IT departments and registries

IT departments and other providers of e-assessment systems must protect the e-assessment system and its database from internal and external threats such as hacking and viruses. Assessment algorithms should remain on the host server. Only test items and the outputs from report generators usually should appear on participants' screens.

Access to test materials should be restricted to qualified and authorised authors, administrators and participants, for example by password or biometric authentication. Invigilation of credit-bearing tests should include identity checking.

IT Departments should verify that the e-assessment system has features to protect it from illegal hacking and computer viruses. Registries should be able to confirm for that reasonable steps have been taken to prevent unauthorised access to servers and assessment materials, which usually means identity checking and invigilation.

For Internet testing, maintain control over the sensitive features of the test and report copyright violations on the Internet. Monitor the web for illegal versions, old/outdated versions and part versions of the Internet test and take steps (e.g., enforcing copyright law) to eliminate these violations.

Take steps to secure protection of test content under existing laws and take appropriate measures to identify stolen test material on the Internet - and to estimate its impact of its distribution on the testing program.

Take appropriate measures to control the distribution of stolen test material on the Internet including notification of appropriate legal authorities. Maintain a process for the adjudication of security breach allegations and specify appropriate sanctions.

(b) Security of data transferred over networks

IT departments

When designing e-assessment systems, build in features that safeguard test-taker data and maintain the security of test material transferred over the Internet. Transactions should be embedded within secure socket layers.

e-assessment systems should enable users to access, check, and/or delete data from the server in accordance with local data protection and privacy legislation. They should have features that ensure regular and frequent backups of all collected data and that allow for recovery of data when problems emerge.

IT departments must maintain the security of test-taker data transmitted over the Internet, e.g., by appropriate levels of encryption. They should ensure that authors in academic departments and participants are informed that the host server has correctly received their data.

IT departments must conduct regular and frequent backups of all collected data and provide users with a detailed disaster recovery plan should problems emerge.

Academic departments and registries

Users should be informed of their rights and obligations in relation to local data protection and privacy legislation. Before e-assessments are delivered over the Internet, they should know the security procedures used to safeguard data transmitted and inform participants thereof.

Confirm with the service provider that they frequently back up data and verify that they are able to allow users to discharge their responsibilities as data controllers under local data protection and privacy legislation such as the EU Directive on Data Protection.

(c) Security of results

IT departments

IT departments must provide secure storage of assessment data. They must maintain the integrity of assessment data by providing technology that does not allow unauthorised altering of information and that can detect unauthorised changes to information. Encryption and password protection should be used to restrict access to assessment data.

Access to personal data stored on the servers must be restricted to qualified and authorised users. Sensitive personal material held on computer, disk, or a server must be Protected with robust (non-trivial) encryption devices or passwords.

Confirm the security and confidentiality of the backup data when used to store sensitive personal data. Know how confidentiality will be maintained when data are stored electronically. Adhere to country-specific data protection laws/regulations governing the collection, use, storage and security of personal data. Protect all material via the use of encryption or passwords when storing sensitive personal data electronically on test centre facilities. Apply the same levels of security and confidentiality to backup data as to the data on the live system when backups are used to store personal data.

Academic departments and registries

When test data must be stored with publishers, procedures and systems to maintain the confidentiality and security of data must be specified. Users must be informed of who has access to test data, for what purposes, and how long it will be stored electronically. Country-specific data protection laws/regulations governing the storage of personal data must be adhered to.

Appendix E: Accessibility

E. A. Draffan

Table 8: Key Points to Look for in Online Assessment

Criterion	Is text re-sizeable?
Why it Matters	People with poor sight may need larger text sizes. People with tunnel vision may need smaller text sizes.
How to check it	Using Internet Explorer, select View > Text size and select either larger or smaller text size
Criterion	Can the active elements of the quiz (eg selection buttons) be selected using the keyboard only?
Why it Matters	Some users find difficulty using a mouse.
How to check it	Click on the Tab key. Successive clicks should highlight the different elements of the page in turn. When an option has been highlighted, pressing the Return key should activate the selection.
Criterion	Do the active elements come up in a logical sequence when selected using the keyboard only?
Why it Matters	Users relying on screen readers use the Tab key to navigate through options. If the options present in a confusing sequence it can be difficult to make sense of the assessment.
How to check it	Click on the Tab key. Successive clicks should highlight the different elements of the page in turn. See if these come up in logical order from top to bottom and left to right.
Criterion	Do selectable options (e.g. radio buttons or check boxes) appear before after the descriptive text?
Why it Matters	Users relying on screen readers want to select an option AFTER they have had it read to them, not before.
How to check it	Inspect the page layout to see if the selection buttons are before or after the text.
Criterion	Where images are used, do they have appropriate descriptions that inform the user without giving away the answers?

Why it Matters	Users relying on screen readers may not be able to see the image so a text description is required. However, the description must not give away the answer - for example the question "Which of the following images represents butane?" should not have the description "Butane molecule" as an [Alt] tag on the image but rather a description of the number and nature of atomic bonds
How to check it	Hover the mouse cursor over the image and see if a screen tip description comes up.

Keyboard navigation relates to the use of Internet Explorer; other browsers may differ.

These guidelines have been taken from JISC TechDis http://www.techdis.ac.uk/resources/sites/staffpacks/Staff%20Packs/E-Assessment/Info%20-%20Online%20Assessment.xml

Also see the W3C standards WCAG 1.0 and WCAG 2.0[21].

References

Ball, S. (2005). A Checklist For Inclusive Assessment: Getting Started With Accessibility (PDF Format - 41Kb). *Proceedings of the 9th International Computer-Assisted Assessment Conference*, Loughborough, UK. http://www.caaconference.com/pastConferences/2005/proceedings/BallS.pdf

Ball, S. (2006). Accessible e-Assessment. CETIS-TechDis Accessibility SIG meeting http://zope.cetis.ac.uk/members/accessibility/meetings/2006/sig14/simonhtml#aims

Ball, S., and Wiles, K. (2002). Accessible E-Assessment: Time For Guidelines? (Word Format - 853Kb). *Proceedings of the Association for Learning Technology Conference 2002*, University of Sunderland, UK. http://www.lifesci.sussex.ac.uk/home/Paolo_Oprandi/projects/alt-c-2002/abstracts2.doc

Thomas, R. C., & Milligan, C. D. (2003). Online assessment of practical experiments. *Proceedings of the Seventh International Computer Assisted Assessment Conference* (pp. 421–430). Leicestershire, United Kingdom, Loughborough University. http://www.jelsim.org/resources/thomasandmilligan_caa2003.pdf

Wiles, K., and Ball, S. (2003). Constructing Accessible Computer-Based Assessment: Minor Works Or Major Renovations? (PDF Format - 886Kb). *Proceedings of the 7th International Computer-Assisted Assessment Conference*, Loughborough, UK. http://www.caaconference.com/pastConferences/2003/procedings/wiles.pdf

[21] http://www.w3.org/TR/WAI-WEBCONTENT/ and http://www.w3.org/TR/WCAG20/

Appendix F: Capability Maturity Models

The assessment maturity model[22] is a recent community development initiated by Question*mark* which identifies four levels of institutional capability. Further details can be seen at the reference linked to.

As an alternative and complementary maturity model, the following discussion[23] suggests five levels of capability (following the original CMM) of an e-learning environment to effectively translate a set of e-learning objectives into an e-assessment.

Level 1: Undifferentiated

Level 1 capability is undifferentiated, and could be called the "Chaos" level. At this level of capability, the process which takes an e-learning intention and produces some e-assessment materials is effectively unknown, unsystematic, unpredictable, and undefined. Level 1 capability probably characterises the efforts of first-time authors and first-time institutional deployment. Of course, the results of level 1 capability might be perfectly adequate; but this would be by accident rather than by design.

Figure 13: "Undifferentiated": Level 1 capability

Level 2: Repeatable

Level 2 capability is called "Repeatable", with the introduction of project management. At this level of capability, the process that takes an e-learning intention and produces some e-assessment materials can be generally repeated and the outcome is somewhat more predictable. This is not to say that the outcome is necessarily of good quality; a repeatable process may simply repeatedly produce poor quality products. Nevertheless, a process which can be reliably repeated is the

[22] https://assessmentmaturitymodel.wikispaces.com/
[23] Based on Gilbert & Gale (2008), *Principles of E-Learning Systems Engineering.*

basis of a process which can be made to produce quality products; the reverse is not true.

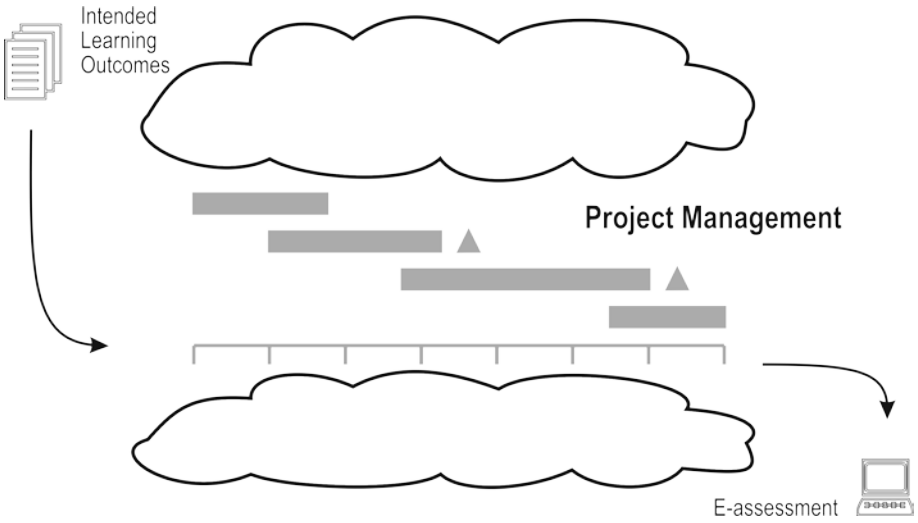

Figure 14: "Repeatable": Level 2 capability

The key capabilities of level 2 process are listed below, Table 9. The focus is on the management and planning of the e-assessment as a whole. Institutions exhibiting level 2 capability may be thought of as showing good intentions, particularly on the project management side, but are still unsure of the correct technical components of e-learning systems development.

ILOs Management	The ILOs baselined & controlled.
	Plans, products, & activities consistent with ILOs.
Managing Materials	Configuration management planned.
	The e-assessment materials identified & controlled.
	Changes controlled.
	Baseline status & content disseminated.
Project Planning	Estimates, activities, & commitments documented.
	Stakeholder commitments agreed.
Project Tracking	Actuals tracked against plan.
	Corrective actions taken to closure.
	Changes agreed by stakeholders.
Subcontract Management	Qualified subcontractors selected.
	Commitments agreed.
	Ongoing communications maintained.
	Prime contractor tracks subcontractor performance and actuals.

Table 9: Level 2 "Repeatable" components

Level 3: Managed

The capability at level 3 is called “Managed”, where standards are introduced, relating to the e-assessment process itself and the ways in which the standardised process can be modified to cater for different kinds of e-assessments. At this level of capability, the process that takes an e-learning intention and produces some e-assessment materials is under active project and product quality management. The process is generally repeatable and predictable, and additionally the standards and product quality assurance activities ensure the production of good quality products.

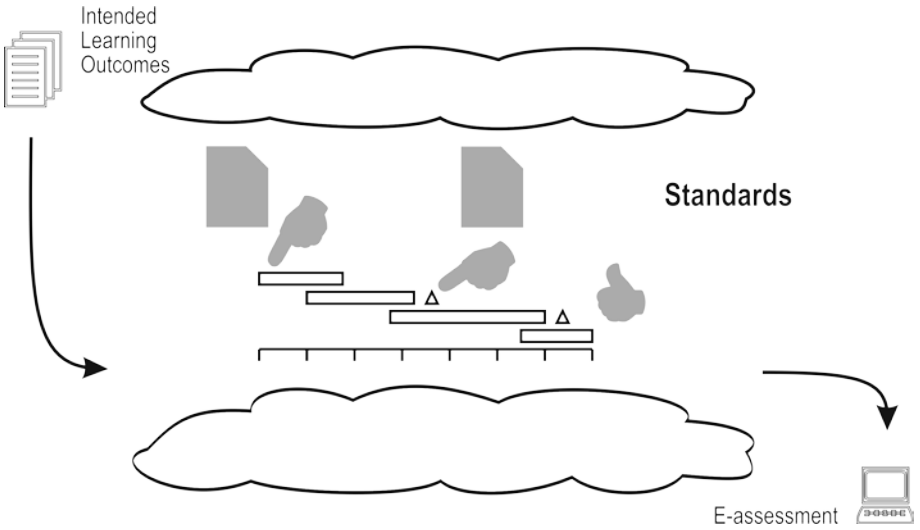

Figure 15: “Managed”: Level 3 capability

The following table, Table 10 lists the level 3 capabilities in more detail. Institutions operating at level 3 capability are operating an e-assessment development process that is technically as well as managerially well-specified.

Process Focus	E-assessment process development and improvement activities planned & coordinated across organization. Process strengths & weaknesses identified.
Process Definition	Standard e-assessment process developed & maintained. Information on the use of the standard process is collected & disseminated.
Integrated Software Management	Particular e-assessment project's process is tailored from organization’s standard e-assessment process. Project is planned & managed according to its defined process.
Product Engineering	Tasks are defined, integrated, consistently performed. E-assessment products are consistent with each other.

Coordination	All stakeholders agree requirements and commitments. Intergroup issues are identified, tracked, and resolved.
Quality Assurance	QA activities planned. Adherence to standards verified. QA results disseminated. Non-compliance addressed by senior management.
Reviews	Peer reviews are planned. Defects in e-assessment products are identified and removed.
Training	Training is planned, provided, & taken by those who need it.

Table 10: "Managed": Level 3 components

Level 4: Quantified

At level 4, the e-assessment process is called "Quantified". Here, the project and quality management processes of e-assessment are quantified and measured. Quantifying, measuring, and managing projects using measurements is known as metrics. The key idea at level 4 is that all of the good practices of levels 2 and 3 are subjected to measurement so that improvement can be managed: numerical targets can be set, the process can be controlled numerically, and the effects of changes to the process can be measured and statistically analysed. In the same way that the e-assessment process could not deliver quality products unless effective project management was in place, the e-assessment process cannot be reliably improved unless metrics are in place and the improvement activities use the metrics to measure and control the required changes.

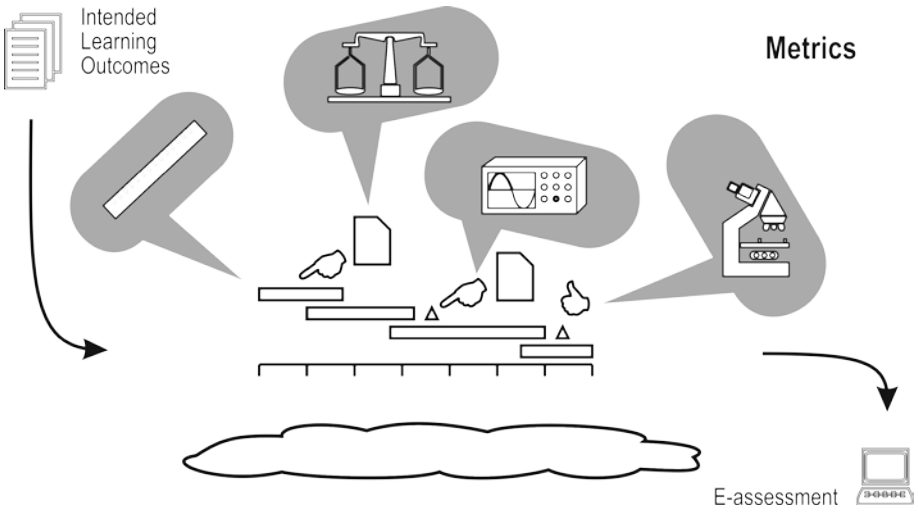

Figure 16: "Quantified": Level 4 capability

The capabilities of a level 4 e-assessment process are shown below, Table 11. The emphasis here is on the quantification of project and quality management procedures and standards to permit managed improvement in these processes and standards. Institutions operating an e-assessment process of level 4 capability are able to estimate with considerable accuracy, for example, the man-days needed to produce an e-assessment, the number of defects it would exhibit on its deployment, and the costs of increasing inspection and piloting to reduce such defects by 50%.

Quantitative Process Management	Quantitative process management activities planned. E-assessment project's process performance controlled quantitatively. Organization's standard e-assessment process capability is quantified.
Quality Management	Quality management activities are planned. Measurable e-assessment quality goals defined & prioritised. Actual progress on achieving goals is quantified and managed.

Table 11: Level 4 "Quantified" components

Level 5: Optimised

The highest level of e-assessment process capability is called "Optimised". Based on the quantification of the e-assessment process achieved in level 4, the capability of the level 5 process is to undertake quality and process optimisation. While the various stages and steps of the e-assessment process were fine-tuned and incrementally improved at level 4, at level 5 this tuning and improvement optimises production for a variety of different kinds of e-assessment project, so the result resembles an engineering "factory" for the efficient and optimum production of high quality e-assessments.

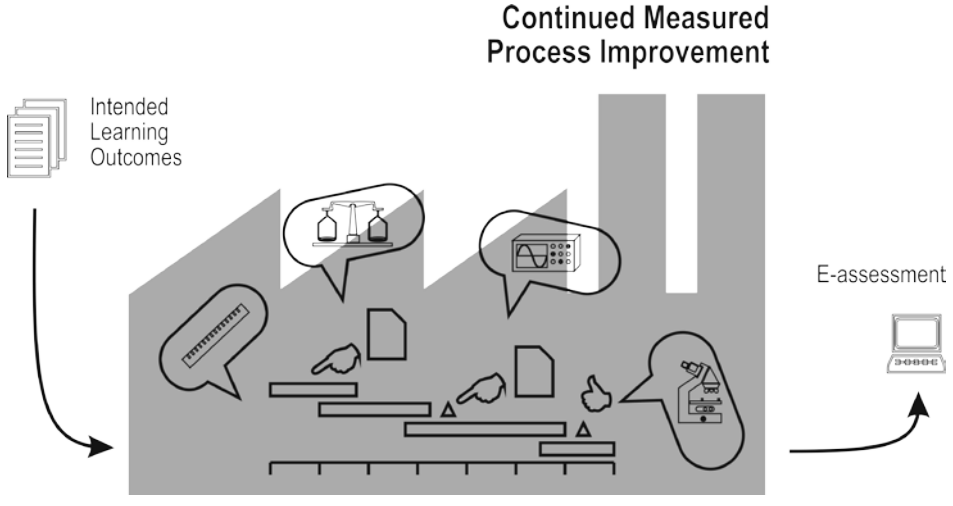

Figure 17: "Optimised": Level 5 capability

Institutions operating a level 5 e-assessment process are operating a process that is under continuous improvement, where the effectiveness of changes to the process can be estimated and demonstrated, and where process changes are optimised. The detailed capabilities of a level 5 process are listed below.

Defect Prevention	Defect prevention activities planned. Common causes of defects sought out and identified. Causes of defects are prioritised and systematically eliminated.
Technology Change Management	Incorporation of technology changes planned. New technology impact on quality & productivity evaluated. Appropriate new technology transferred into normal practice.
Process Change Management	Continuous process improvement planned. Organization-wide participation in process improvement. Organization's standard e-assessment process & projects' defined processes continuously improved.

Table 12: "Optimised": Level 5 components

Appendix G: QAA (1994) guidelines on Student assessment

The QAA (1994) guidelines can be found online.[24]

The following are extracts from the section on Assessment.

60 Assessment has many purposes. According to the CNAA, the purpose of assessment is to enable students to demonstrate to examiners the extent to which they have fulfilled the objectives of their programme of study. In the consultation, institutions pointed to other purposes of assessment. For example, assessment is an integral part of students' learning experience affecting their motivation and providing them with insight into the extent they have fulfilled objectives and their capability to utilise their own learning. Assessment was also seen as a teaching tool requiring students to marshal information and present their thoughts clearly.

61 The CNAA recommended that each programme of study leading to an award should have a Board of Examiners and that in normal circumstances no student should be a member of the Board or attend its meetings. CNAA required that detailed and accurate written records of the proceedings of each Board of Examiners be maintained.

62 Some of the issues raised by CNAA and CVCP that need to be considered with regard to assessment and the programme of study include:

62.1 whether the assessment relates to the learning objectives/outcomes of the programme;

62.2 whether the assessment methods reflect the teaching strategies used in the programme;

62.3 the objectives of the assessment and the criteria for success;

62.4 whether there is fairness and consistency in marking schemes for all forms of assessment such as in-course assessment, formal examination, project work;

62.5 the means to determine the classification of awards; and

62.6 whether assessment methods are consistent with the requirements of such external accreditation agencies as may be appropriate.

63 With regard to assessment and the needs of students, the following are seen to require consideration by institutions:

63.1 whether students are aware of the nature, methods, relative weighting and timetable of assessment during and at the end of a programme;

63.2 whether students are aware of the consequences of assessment; and

63.3 whether the scheduling of assessed assignments avoids excessive work overload (this may be a particular problem in respect of modular programmes).

64 The types of procedural issues relating to assessment that require consideration by institutions include:

64.1 procedures for the conduct of all assessments;

64.2 procedures for the conduct of Boards of Examiners or other equivalent bodies;

64.3 arrangements for involving external examiners in assessment (paragraphs 69 to 77);

[24] http://www.qaa.ac.uk/search/publications/archive/DQE100_GuidelinesQualityAssurance 1994.asp

64.4 guidance for examiners and students on how to deal with evidence of circumstances that may have affected student performance;

64.5 procedures on how to deal with academic misdemeanours; and

64.6 procedures for exclusion of a student, if appropriate.

In assessing students, institutions will wish to ensure that they have taken into account issues relating to programmes of study and to the needs of students, and to operate appropriate procedures for the conduct of assessments.

65 The nature of postgraduate research awards means that institutions need procedures that relate directly to the examination of these students. For example, students will need information on the procedures for the nomination of examiners and on the examination process. Both the CVCP and CNAA recommended that there should always be at least two examiners of whom at least one should be external to the institution. The student's supervisor should be the internal examiner only in exceptional circumstances, and in this event an additional examiner (internal or external) should also be appointed.

66 The CVCP and CNAA made similar recommendations on the examination of postgraduate research students. Once the examiners have been appointed they should read the student's thesis and prepare independent preliminary reports. Following the oral examination of the student, they should prepare a joint report with a recommendation for one of a series of options which might include: the immediate award of a degree; award after minor amendments made to the satisfaction of the internal examiner; revision and re-submission; formal submission for a lower degree with or without amendments; or rejection and no right of re-submission.

Checklist for quality assurance systems

In this section, all the guidance cited throughout the document has been drawn together in one place to provide a useful checklist for institutions. The checklist should help those that may wish to compare their present arrangements with the guidance provided throughout the document.

Establishing a framework for quality

1 An effective quality assurance and control system is characterised by agreement throughout an institution on purposes and methods and includes a feedback loop to inform and improve the quality of educational provision.

2 An effective quality assurance and control system is underpinned by wide participation, effective channels of communication, the collection of acceptable evidence, the acceptance of responsibility by staff and students, and an institutional commitment to staff development and training.

Reviewing quality assurance systems

3 Institutions will wish to ensure that they review the operation and effectiveness of their quality assurance and control procedures and that the means adopted reflect the particular approach to quality of each institution.

Admissions policies

4 Institutions will wish to address the following issues in their admissions procedures:

4.1 the provision for equality of opportunity for all applicants;

4.2 the reasonable expectation that anyone admitted to a programme of study will be able to fulfil the objectives of the programme and achieve the standard required;

4.3 the provision of evidence by applicants of relevant personal, professional and educational experiences; and

4.4 whether it might reasonably be expected that a student's proposed programme of work can be completed in the designated timescale and whether proper supervision can be provided and maintained.

Admission requirements

5 Institutions will wish to provide clear and accurate information on all the available admission routes and on any associated requirements for entry to programmes of study.

Information for prospective students

6 Institutions will wish to provide a range of information that meets the needs of prospective students and they will wish to have procedures to ensure that the information provided is useful, accurate and realistic.

7 The marketing strategy of an institution of higher education needs to be appropriate to its mission and directed towards its potential clients.

Pre-entry guidance

8 Institutions will wish to ensure that appropriate structures and provision operate to offer pre-entry guidance and support for prospective students.

The selection process

9 In the selection process, an institution will wish to:

9.1 make the details of its selection process clear to applicants;

9.2 ensure that staff responsible for admissions are aware of entry policy and criteria;

9.3 ensure that admissions staff are conversant with the programme syllabus and available options;

9.4 provide guidance to admissions staff on which students should be interviewed;

9.5 ensure that the selection procedure used operates fairly for all types of applicants, regardless of their background;

9.6 ensure that staff are suitably trained to select all types of applicants and to make fair and sound judgements having regard to the admission criteria; and

9.7 ensure that satisfactory procedures are established for selecting applicants and that procedures are consonant with institutional policy.

Facilitating student entry

10 Institutions will wish to ensure that they operate quality control mechanisms to monitor their admission and selection policies and procedures and the accuracy of their promotional materials. There should be a feedback loop to review existing practice.

Quality assurance and the diversity of higher education

11 An institution will wish to ensure that all approaches to teaching/ learning are scrutinised appropriately through its quality assurance and control systems in order to ensure the comparability of standards and the quality of the student experience.

12 An institution will wish to ensure that it has appropriate procedures to assure efficiently the quality of learning within individual modules and also to assure the coherence and progression of learning within individual programmes.

External programme approval

13 Institutions will wish to offer guidance and support to staff regarding the external accreditation available for programmes, the advantages of securing such accreditation, and the procedures and criteria to obtain accreditation.

Internal programme approval

14 When approving new academic programmes or modular frameworks, or revising existing programmes, institutions will wish to have quality assurance systems that reflect the structure of academic provision and maintain the quality of programmes and the standards of awards.

15 Institutions will wish to devise appropriate structures and procedures for the quality assurance of credit from prior learning, work-based and other experiential learning, together with mechanisms for assuring coherence within individual awards.

16 Institutions will wish to state the nature of the information which should be included in new programme proposals and how such information is to be used by those charged with responsibility for developing and approving programmes.

17 In formal procedures to scrutinise programme proposals, institutions will wish to include:

17.1 details on who will review the proposal, with external representation as appropriate;

17.2 a framework for considering the programme proposal to include a mechanism to determine the academic credibility and resource implications of the proposal;

17.3 arrangements, where appropriate, for external professional body interest to be included; and

17.4 details of the relationship and requirements of the Senate or Academic Board in the approval process.

Programme information for students

18 Institutions will wish to consider producing information for students on their chosen programme of study.

Teaching and learning

19 Institutions will wish to consider how different teaching strategies bring about their intended student learning objectives and enable students to take maximum responsibility for their own learning.

Evaluation of programmes of study

20 An institution will wish to encourage the adoption of a variety of ways to ensure that programmes of study are running as planned. These might include *ad hoc* day-to-day checks on details, annual monitoring and formal review processes. In the evaluation of programmes of study, the following points need to be taken into account:

20.1 the organization and location of the programme (eg, modular programme, work-based/ franchised programme);

20.2 the frequency of evaluations and who conducts them;

20.3 the relationship to external (eg, professional bodies) review procedures;

20.4 a mechanism to review or monitor programmes (using the information in the listing above);

20.5 what happens as a result of the evaluation (action taken following preparation of monitoring reports or full programme reviews); and

20.6 a mechanism for final endorsement of the review (eg, the role of the Senate or Academic Board).

21 Institutions will wish to have in place clear mechanisms for deciding upon and effecting the major modification or cessation of a programme, which include safeguarding the interests of the students already in or about to join the programme.

Evaluation of teaching and learning

22 Institutions will wish to consider how best to review the effectiveness of teaching and student learning, and where appropriate, to develop further approaches to evaluation.

Staff appointment

23 Institutions will wish to ensure that their appointment procedures take into account the competence and aptitude of staff with regard to the full requirements of the position.

Staff development and training

24 Institutions will wish to promote systems and procedures for quality enhancement to encourage and sustain a culture in which all students and staff contribute to a process of continuing improvement.

Staff appraisal

25 Institutions will wish to have an appraisal system in place for all staff. An appraisal system needs to:

25.1 be explicit regarding its role in relation to staff development and career progression;

25.2 be confidential and supportive;

25.3 specify the procedures, materials and information that will be used in the appraisal;

25.4 specify all the possible outcomes of the appraisal process;

25.5 identify who will conduct the appraisal;

25.6 provide appropriate training for appraisers and appraisees; and

25.7 recognise the contribution of teaching for academic staff in the appraisal.

Quality of collaborative arrangements

26 When entering into an academic collaborative arrangement, institutions will wish to produce a formal, written statement that sets out the responsibilities and duties of each of the participating institutions in respect of the maintenance of standards and the protection of the student experience.

Postgraduate research students

27 Institutions of higher education will wish to have guidelines on postgraduate research students that are known to students and supervisors. These guidelines will normally include information on:

27.1 procedures for the appointment of one or more supervisor(s) who are suitably qualified for the proposed subject;

27.2 specification of roles and responsibilities of the supervisor(s) and the student, especially with regard to guidance, extent of contact and progress;

27.3 guidance on the resources and facilities available to postgraduate research students (eg, use of photocopiers, computing equipment, common room);

27.4 guidance to supervisors and students on reporting arrangements and requirements;

27.5 procedures and requirements for conversion from master's degree to doctorate;

27.6 procedures for the change of a supervisor;

27.7 grievance and appeals procedures;

27.8 details of training available for supervisors; and

27.9 mechanisms for monitoring and ensuring that the various faculty and departmental arrangements for postgraduate research students are in accordance with policy.

28 Institutions of higher education will wish to provide training for postgraduate research students appropriate to their needs.

Student support services

29 Institutions will wish to offer a range of student support services appropriate to the needs of students, and to establish quality assurance and control systems to ensure the suitability and effectiveness of these services.

Student grievance

30 Institutions will wish to operate widely known and understood policies and procedures to deal with student complaints.

Student progress

31 Institutions will wish to have procedures in place to ensure that students receive regular, frequent and prompt feedback on their progress and performance in relation to their chosen programme of study.

Student assessment

32 In assessing students, institutions will wish to ensure that they have taken into account issues relating to programmes of study and to the needs of students, and to operate appropriate procedures for the conduct of assessments.

33 Institutions will wish to ensure that postgraduate research students are informed of examination procedures, including the nomination of examiners, the examination process and the possible outcomes.

Appeals

34 Institutions will wish to have clear, formal and well publicised procedures on appeals for all students, including those who study 'off-site' or are on 'collaborative' programmes.

External examiners

35 Institutions will wish to ensure that they operate an effective system for external examiners and pay particular attention to the selection and appointment of external examiners and the nomination of moderators; contractual arrangements; the role of external examiners; the form of the external examiner reports; and the arrangements for review and implementation of the external examiner recommendations.

Definitions of quality

British Standards Institution (from BS4778)

Quality. The totality of features or characteristics of a product or service that bear on its ability to satisfy a given need.

Quality control. The operational techniques and activities that are used to fulfil requirements for quality.

Quality assurance. All those planned and systematic activities to provide adequate confidence that a product or service will satisfy given requirements for quality.

Quality audit. A systematic and independent examination to determine whether quality activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve objectives.

Division of Quality Audit, Higher Education Quality Council

Quality control is defined as an operational function, applied at all levels by an institution to the management of its activities related to teaching and learning. It is ITC Computer-based and Internet delivered testing Guidelines (2005)

concerned with the checks and measures by which a body determines during, or shortly upon completion, that the operations for which it is responsible are working as planned and intended, including their fitness for the purpose specified and that resources have been optimised and identified goals achieved.

Quality assurance is defined as the arrangements by which an institution discharges its corporate responsibility for the quality of the teaching and learning it offers by satisfying itself that its structures and mechanisms for monitoring its quality control procedures are effective and, where appropriate, that they promote the enhancement of quality.

Quality audit. This is defined in identical terms to those used in British Standard 4778 for 'Quality Audit'.

Higher and Further Education Act 1992 and the White Paper: 'Higher Education: A New Framework 1991'

Quality control relates to the arrangements (procedures, standards, organization) within HE institutions which verify that teaching and assessment are carried out in a satisfactory manner. Quality control would include the external examiner system but would usually be post hoc and the responsibility of the institution itself.

Quality assurance encompasses all the policies, systems and processes directed to ensuring maintenance and enhancement of the quality of educational provision in higher education.

Quality audit is the process of ensuring that the quality control arrangements are satisfactory. In practice, prime responsibility for quality audit lies individually or collectively with institutions, and it extends to the totality of quality assurance in an institution including staff development and curriculum design. External quality audit will be conducted by the Quality Audit Division of the Higher Education Quality Council which collectively owned and funded by the institutions.

Appendix H: E-assessment policies

John Winkley

Summary

In current practice, e-assessment generally makes use of a discrete set of technologies to perform a specific purpose and is reasonably widely used in JISC's HE and FE communities. However, the regulatory and policy framework which surround the use of e-assessment is complicated because:

- In the FE sector, summative e-assessment is covered by the external regulation of UK public examinations. Colleges need to follow guidance and requirements provided by the UK regulators and their awarding bodies.
- In the HE sector, universities regulate their own examinations, and so have individual assessment policies which may include rules covering the use of e-assessment. QAA also provide some guidance in this respect.
- In all sectors, the use of formative e-assessment is largely covered by policy and practice relating to the use of ICT in learning – for example, e-learning and ILT strategies, which are published by a range of organisations including most notably JISC and Becta. Institutions may have their own policies too.

This leads to a complicated array of guidance, regulations and recommendations which this document presents. While the regulatory arrangements for any particular situation are generally clear and unambiguous, they tend to draw on a long list of nested documentation, and differ from setting to setting. It is not really practical for anyone other than a specialist to be fully aware of the requirements across the entire range of settings. It is also unlikely that the regulatory arrangements will become simplified in the near future because of the different and well-established educational administrative structures that exist in JISC's different communities.

On a positive note, unsurprisingly, much of the varied documentation, is, in fact, very similar in purpose and approach. For example, in most cases, the variation to examination practices when moving from paper to on-screen is limited to only those matters which must be changed (for example, special access requirements). Otherwise, the approaches are as identical as possible. This means in practice that providers may be confident that working from their existing familiar institutional assessment and ILT/e-learning policy documents as a baseline, and considering amends to practice only where necessary, is likely to be a satisfactory approach.

Policy and Regulation in UK e-assessment

The regulation and policy relating to e-assessment in the UK is divided according to sector and purpose as follows:

Sector

- In **Further Education** (and schools) almost all qualifications are awarded by external "Awarding Bodies" (also known as Exam Boards). Their conduct is regulated by national Regulators, Ofqual in England (previously QCA), DCELLS

in Wales, SQA in Scotland, CCEA in Northern Ireland. The English regulator is fully independent from the awarding bodies. In Wales, the regulator is part of the Government Education department. In Scotland and Northern Ireland, the regulator is also the dominant awarding body. The regulatory arrangements vary for different groups of qualifications: there are different arrangements for general qualifications (Diploma, GCSE and A Level) and vocational qualifications (BTEC, NVQ, etc). The regulatory arrangements for e-assessment tend to be additional to the arrangements for the qualification.

- In **Higher Education**, universities award qualifications themselves, set their own policies, and largely self-regulate. QAA provide a light touch public-interest regulation funded by subscriptions from HEIs. QAA's website makes only limited reference to e-assessment, but states that regulatory principles for assessment should be applied. Regulatory and policy documentation in the HE sector is much more limited than the school/college sectors.

Purpose

- **Assessment for Accreditation**: where the results of the assessment contribute towards an award (for example, a qualification or a unit) the regulatory organisations (described above) take a much greater interest in the assessment arrangements.
- **Assessment for other purposes**: the majority of assessment is undertaken within learning programmes for a variety of purposes including most commonly, placement, diagnostic, formative and programme monitoring activities. While many general regulations apply in these settings, the majority of the specific "examination" regulations do not apply.

Regulation Documentation for e-assessment

Important notes:

The regulatory arrangements for e-assessment are evolving (often alongside or retrospective to developing practice). As a result it is likely that this document will become out of date. Readers are advised to check with the authority websites for updates.

There is substantial collaboration between the four nations and between sectors for developing and agreeing regulatory approaches and associated documentation – this is reflected in much of the documentation.

In practice, e-assessment is part of a wider agenda for e-learning. For brevity, this document does not include the wider policy and strategy documents relating to the development of e-learning. See https://wiki.jisc.ac.uk/display/elearning/Strategy+and+policy for this wider set.

Policy and Guidance Documents in Higher Education

The JISC e-learning programme homepage, including details of the JISC e-assessment priority activity plan.	JISC	http://www.jisc.ac.uk/whatwedo/themes/elearning/programmeelearning.aspx
Guide to Effective Practice with e-Assessment	JISC	http://www.jisc.ac.uk/publications/publications/pub_eassesspracticeguide.aspx
REAP (Re-engineering Assessment Practices) reports on the redesign of formative assessment and feedback practices across three pilot institutions.	Scottish Funding Council	http://www.reap.ac.uk/
Part of the JISC Infonet "Effective Use of VLEs" resource relating to e-assessment	JISC infonet	http://www.jiscinfonet.ac.uk/InfoKits/effective-use-of-VLEs/e-assessment/index_html
National Student Survey of Satisfaction		http://www.thestudentsurvey.com/

National Student Survey extract	**2007**	**2008**			
Question	**Registered HEI**	**Registered HEI**	**Taught HEI**	**Taught FEC**	**Com-bined**
Assessment and feedback	%	%	%	%	%
5 - The criteria used in marking have been clear in advance.	68	69	69	74	70
6 - Assessment arrangements and marking have been fair.	71	72	72	76	72
7 - Feedback on my work has been prompt.	53	56	56	59	56
8 - I have received detailed comments on my work.	58	61	61	73	61
9 - Feedback on my work has helped me clarify things I did not understand.	53	56	55	68	56

The scores of 50-60% are the lowest in the whole survey.

Regulatory Documents for e-assessment in Further Education and Schools

Open letter clarifying intentions for a single technology platform for e-assessment in England.	Ofqual	http://www.ofqual.org.uk/files/Open_letter_technology_platforms.pdf
The regulatory principles for e-assessment (April 2007) agreed by the regulators of England, Wales, Northern Ireland and Scotland.	Ofqual, SQA, CCEA, DCELLS	http://www.ofqual.gov.uk/files/Final_regulatory_principles_document_-_PRINTED.pdf
The e-futures website designed to encourage and support the use of e-assessment in vocational qualifications.	Ofqual	http://www.efutures.org/graphics/index.html
The e-Assessment Guide to effective practice. A resource for those involved in the management and administration of e-assessment in awarding body centres where accredited qualifications are delivered.	QCA, SQA, CCEA, DCELLS	http://www.qca.org.uk/libraryAssets/media/qca-07-3209_e-Assessment_Guide.pdf (QCA, CCEA, DCELLS English version)http://www.sqa.org.uk/sqa/23369.html (SQA version)
SQA Guidelines on Online Assessment for Further Education (2003)	SQA	http://www.sqa.org.uk/files_ccc/GuidelinesForOnlineAssessment(Web).pdf
SQA's Vision and Strategy for E-Assessment (2007)	SQA	http://www.sqa.org.uk/files_ccc/VisionandStrategyforE-assessment_June07.pdf
SQA Guidelines on e-assessment for Schools (2005)	SQA	http://www.sqa.org.uk/files_ccc/SQA_Guidelines_on_e-assessment_Schools_June05.pdf

Regulatory documents for external qualifications assessment in schools and colleges

The statutory regulation of external qualifications in England, Wales and Northern Ireland (2004)	Ofqual, DCELLS, CCEA	http://www.ofqual.gov.uk/321.aspx
GCSE, GCE, GNVQ and AEA code of practice (April 2008)	Ofqual, DCELLS, CCEA	http://www.ofqual.gov.uk/191.aspx
NVQ code of practice (2006)	QCA, DCELLS	http://www.ofqual.gov.uk/1511.aspx

Regulatory arrangements for the Qualifications and Credit Framework (QCF) (August 2008)	Ofqual, CCEA, DCELLS	http://www.ofqual.gov.uk/files/ Regulatory_arrangements_ QCF_August08.pdf
Awarding body criteria (SQA 2007). Details the criteria that SQA's Accrediting Body use to accredit qualifications and to approve and monitor Awarding Bodies	SQA	http://www.sqa.org.uk/files_ccc/ AccreditationAwardingBodyCriteria.pdf
Regulatory monitoring and reporting guidance (2007). Description of the relevant regulators' routine arrangements for monitoring and reporting publicly on accredited qualifications	Ofqual, DCELLS, CCEA	http://www.ofqual.gov.uk/462.aspx

Note that regulatory arrangements for qualifications outside of JISC's interest area (for example testing within the National Curriculum) are not included here.

Other e-assessment regulatory documents

There is an international standard for e-assessment:

BS ISO/IEC 23988:2007 Information technology. A code of practice for the use of information technology (IT) in the delivery of assessments

(http://www.bsi-global.com/en/Shop/Publication-Detail/?pid=000000000030135433)

This was developed in the UK and supersedes a previous British Standard for e-assessment:

BS 7988:2002 Code of practice for the use of information technology (IT) in the delivery of assessments

(http://www.bsi-global.com/en/Shop/Publication-Detail/?pid=000000000030044574)

This standard gives recommendations on the use of information technology (IT) to deliver assessments to candidates and to record and score their responses. Its scope is defined in terms of three dimensions: the types of assessment to which it applies, the stages of the assessment "life cycle" to which it applies and this British Standard's focus on specifically IT aspects.

E-assessment as part of wider e-learning/ICT strategy

Harnessing Technology Strategy. (Harnessing Technology: Next Generation Learning 2008-14: A summary) This is the revised strategy for 2008 and covers developing the use of ICT for schools and FE colleges. There is a summary of the strategy at http://publications.becta.org.uk/display.cfm?resID=37346

There is a web-based presentation of the strategy at: http://www.becta.org.uk/nextgenerationlearning.php

Accessibility

(Links are provided to government pages with introductory text and links to the legislation, rather than the legislation itself)

Accessibility in e-Assessment Guidelines Commissioned by TechDis, following consultation with the E-Assessment Group (membership listed on page 4 of the report), to examine the state of guidance for accessibility in e-assessment in the UK. (The nearest to a common statement of e-assessment accessibility policy/requirements that currently exists)	http://www.techdis.ac.uk/index.php?p=9_1_11
The Disability Discrimination Act 1995 and 2005	http://www.direct.gov.uk/en/DisabledPeople/RightsAndObligations/DisabilityRights/DG_4001068
The Disability Rights Commission Act 1999	http://83.137.212.42/SiteArchive/drc_gb/the_law/legislation__codes__regulation/dda_and_related_statutes.aspx.html
The Special Educational Needs and Disability Act 2001.	http://83.137.212.42/SiteArchive/drc_gb/the_law/legislation__codes__regulation/dda_and_related_statutes.aspx.html
The Human Rights Act 1998	http://www.direct.gov.uk/en/Governmentcitizensandrights/Yourrightsandresponsibilities/DG_4002951

Cheating

Digital technologies and dishonesty in examinations and tests: this report, first published in December 2006, gives an overview of the current state of knowledge about the impact of digital technologies on dishonest practice by students in examinations and tests: http://www.ofqual.gov.uk/177.aspx

Membership Organisations

The E-Assessment Association: “The eAA campaigns for the widening of the effective use of e-Assessment in support of learning. The eAA will do this by encouraging individuals and organisations to make good use of e-Assessment, by explaining the issues to the public, and by influencing educational thinking and policy.” http://www.e-assessmentgroup.net/

The Chartered Institute of Educational Assessors: “The CIEA is a professional body dedicated to supporting the needs of everyone involved in educational assessment. Our members include everyone involved in assessment, from senior examiners, moderators and markers to individuals with an interest in or responsibility for assessment in primary schools, secondary schools, colleges, universities, training centres and other educational organisations.” http://www.ciea.org.uk/

Glossary of terms

For regulation purposes, the FE sector has adopted the JISC Glossary of e-assessment terms: http://www.jisc.ac.uk/assessment.html

Organisation links

Ofqual: www.ofqual.org.uk

CCEA: www.ccea.org.uk

SQA: www.sqa.org.uk

DCELLS: http://new.wales.gov.uk/about/departments/dcells/;jsessionid=JhPQJkrLGCC9NQ1RqttGH3SvJDmK6km9tYqGrfG2HKT9J8sZTZhH!-1414164158?lang=en

QAA: www.qaa.org.uk

QAA Code of Practice for assessment: http://www.qaa.org.uk/academicinfrastructure/codeOfPractice/default.asp

JISC: www.jisc.ac.uk/assessment

Techdis: www.techdis.ac.uk

Appendix I: Trajectory models and Dual Path Theory

Uptake Trajectories in HE

Individual assessments are widely understood to have a three-phase lifecycle: Authoring, Delivery and Reporting. Authored e-assessments are delivered to participants and the interpreted results are used to inform the next iteration of the authoring process, and so on. However, technology-based systems have a lifecycle of their own (Moore, 1999; Rogers, 2003) and Warburton (2008) found that the uptake of e-assessment within UK HEIs appears to be characterized by risk trajectories. According to Warburton's analysis, tutors' individual experiences of e-assessment can be described in terms of individual 'trajectories' that are a fundamental element that underlies bulk institutional uptake. A tutor's e-assessment trajectory differs critically from otherwise similar patterns of technology uptake such as VLE use in that a significant element of risk attends technology-based assessment activities, particularly in credit-bearing assessment.

Individual tutors' e-assessment trajectories are characterised as high or low risk according to the fashion in which they progress towards high stakes assessment. Planned sequential approaches incur only moderate increments of risk at each step, which results in a typically low risk trajectory. Warburton found that the biggest influences on tutor trajectories were to be their motives for using e-assessment: where the aim was primarily to improve learning and teaching practice, the consequence was a sustained progression through the different stages of e-assessment use that resulted in lower risk trajectories, as below.

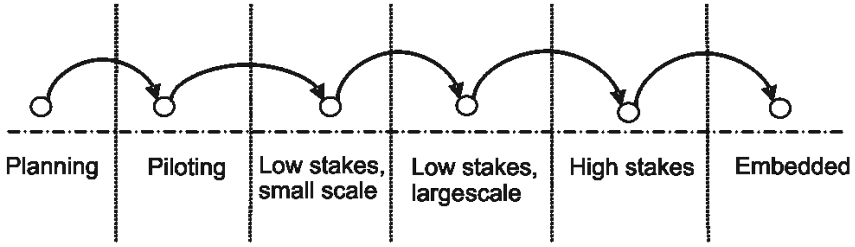

Figure 18: Trajectory model showing incremental institutional uptake (low-risk)

On the other hand, where the aim was primarily to secure productivity gains, an ad hoc style of use resulted in higher-risk trajectories, as below.

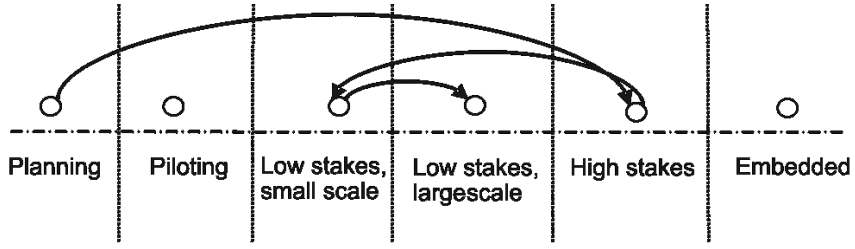

Figure 19: Trajectory model showing ad hoc tutor uptake (higher risk)

The trajectories which emerged from selective coding as most typical of current CAA practice are illustrated below. They are:

Adhering to the structured institutional strategies (linear institutional) incurs least risk and the greatest chance of progression to the technology becoming embedded.

Tutors who stick to a plan (linear) and implement CAA starting with low stakes then moving to progressively higher stakes testing incur least risk.

Leapfrogging stages (ad hoc) entails taking on unnecessary risk which may result in difficulties and the technology being abandoned (disastrous).

The most likely cause of CAA testing being abandoned appears to be an originating tutor moving on and no-one picking up the baton (cut short).

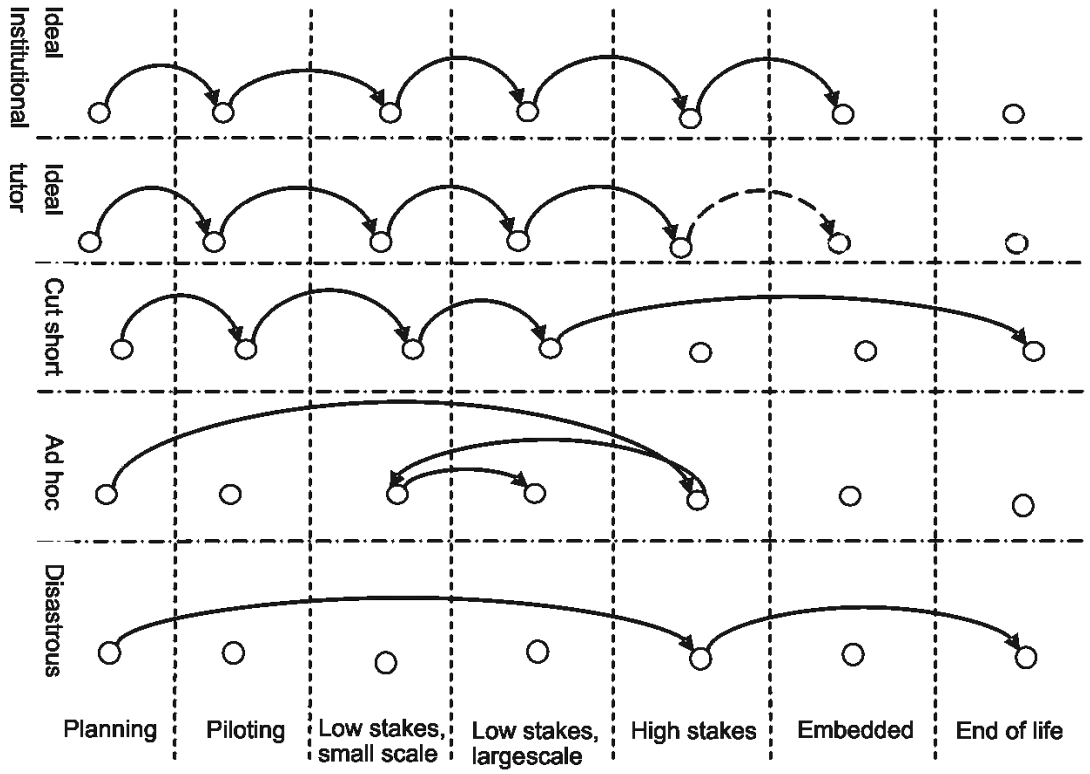

Figure 20: Trajectory model

Dual path theory

The core category: tutors' motives

The pattern of CAA uptake over time at the level of individual tutors - their 'trajectory' - is the fundamental unit which, on the micro scale, underlies institutional uptake on the macro level:

> This only makes sense in bulk when you consider a large number of tutors all going through this cycle and then the aggregate effect of all this will be a trend towards either doing it or not doing it, through these various mechanisms. (Learning technologist LtN2M003)

A tutor's CAA trajectory differs critically from otherwise similar patterns of technology uptake such as VLE use in that a significant element of risk attends technology-based assessment activities, particularly in credit-bearing assessment:

> When you get failures... its very public, rather more than computer-based learning... [but] if email runs down for 10 minutes well, try again later; so the stakes are higher. (Learning technologist LtO5M002)

Individual CAA trajectories can be broadly characterised as high or low risk according to the fashion in which tutors progress towards high stakes assessment. Where uptake proceeds in a planned sequential fashion from testing through formative to low and then high stakes summative testing, small increments of risk are incurred in each step which results typically in a linear low risk trajectory. Where uptake proceeds directly to summative use, large increments of risk may be incurred at once which results typically in a non-linear high risk trajectory. The biggest influences on the CAA trajectory described by a tutor were their motives for using CAA in the first place. Where the aim was primarily to secure productivity gains the consequence was an ad hoc style of use that resulted in high risk trajectories. Where the aim is primarily to improve learning and teaching practice the consequence is a more sustained progression through the different stages of CAA use that results in lower risk trajectories.

Quick wins?

Time pressures on tutors across the sector are well documented (Bull, 1999; Gibbs, Habeshaw and Yorke, 2000) and are often compounded by increasing demand for research output that will raise their profile in the next research assessment exercise (RAE). This promotes a utilitarian approach to assessment activities which prizes quick returns above pedagogic gains or longer term considerations such as an expected reduction in assessment load once a large item bank has been built. CAA was widely acknowledged to offer the potential of productivity gains in terms of more efficient authoring, publication, delivery, marking and reporting, which was summed up by some as an effective reduction in paperwork:

> ...paper based system was very slow and required large resources to keep track of assessment, moderation, results, retries, etc' - Learning technologist LtN4M001).

However it also emerged that where unsupported tutors sought these 'quick wins' without investing in preparative activities such as seeking the advice of experienced colleagues or setting up formative exercises and practice quizzes, the degree of risk taken on all at once could be so significant that colleagues were discouraged from

using CAA themselves. This effect was prominent in extreme cases such as student data was loss during an invigilated CAA failure:

> ... when the email came round about the [CAA] disaster... some of those colleagues... just went non-linear... how can we possibly have... taken on something which under the most fundamentally obvious things that it had to work under, it fails at the first hurdle? (Tutor AmO5M007)

The effect was less pronounced where the unfavourable outcome was limited to unplanned expenditure of time and effort, for example to recover data or reassure students:

> ... this is taking a lot of our time, correcting really quite trivial errors... I've spent a lot of time doing quality checks on in-coming CAA tests. (Learning technologist LtO5M002)

Failure to think through the implications of using CAA can have serious implications:

> ... a CAA had been taken and the results had been distributed to [an inexperienced] tutor, the tutor had given them to someone... who... sent them to an external [examiner], including a detailed breakdown of the item analysis of the assessment, which the tutor didn't understand and hadn't intended to go. So the external [examiner] looked at all this and said 'thank you very much, your test appears to be invalid'. (Learning technologist LtO3M001)

Unintended outcomes of this kind threaten the CAA user's credibility. The increased risk incurred by productivity-driven approaches to CAA applications and the braking effect they have on uptake by colleagues represents an extreme case and is shown in the upper half of the paradigm model (Figure 1). It should be noted that this opening of the assessment process to public scrutiny could be regarded as an unintended consequence of CAA which is seldom included in risk registers. Until recently assessment feedback was rarely given, not least because the examination system was ill equipped to provide it. Therefore participants didn't expect feedback and there was no possibility of a debate about academic standards. Now people know it can be done so they take it for granted, not only for formative and diagnostic use but also for summative assessment as well.

Slow burn?

Conversely, where tutors aimed primarily for pedagogical improvements they incurred much less risk and the resultant trajectories were characterised by routine use of small scale quizzes with an initial emphasis on low stakes testing such as formative and diagnostic applications. This sometimes progressed towards higher stakes testing on a larger scale:

> [Their] work on this... [began with] pilot development stages. We [proceeded] to running some Key and Basic Skills tests... with the prospect of widening this out to regular sessions, on-demand and then to high stakes options. (Learning technologist LtS6M002)

This staged, incremental approach was encouraged and supported by learning technologists who recognised the value for tutors of learning to use complex CAA tools in less critical applications. High stakes applications such as examinations were seen by learning technologists very much as risky undertakings which should be the final goal of CAA trajectories:

> [We] did a needs analysis of who wanted to use it... then my team... decided to pilot formative, summative...They wanted to go from there to using it for

exams because they had already sort of built a little bit of confidence up. (Learning technologist LtN4F001)

Experienced tutors made similar points:

For final summative assessment on-line, there are many technical difficulties to overcome and the whole operation is much more complex than a normal paper exam... [it's got to be] planned for. (Tutor AmO5F001)

Staged lower risk trajectories of this kind generally produced productivity gains and consequently diffusion tended to be steady rather than spectacular. Where tutors emulated this approach, they appeared to do so because they perceived a structured, methodical pattern of practice which would protect their investment in assessment materials and which might yield sustainable if modest productivity gains in the medium to long term:

[We saw that] those that do them get instant feedback on their progress. (Tutor AmO3F001)

[We saw] ...higher motivation of students through immediate feedback. Students persist until they get the right solution... (Tutor AMO5O002)

The reduced risk incurred by pedagogically-driven attitudes to CAA use and the accelerating effect this has on uptake by colleagues is shown in the lower half of the model (Figure 21).

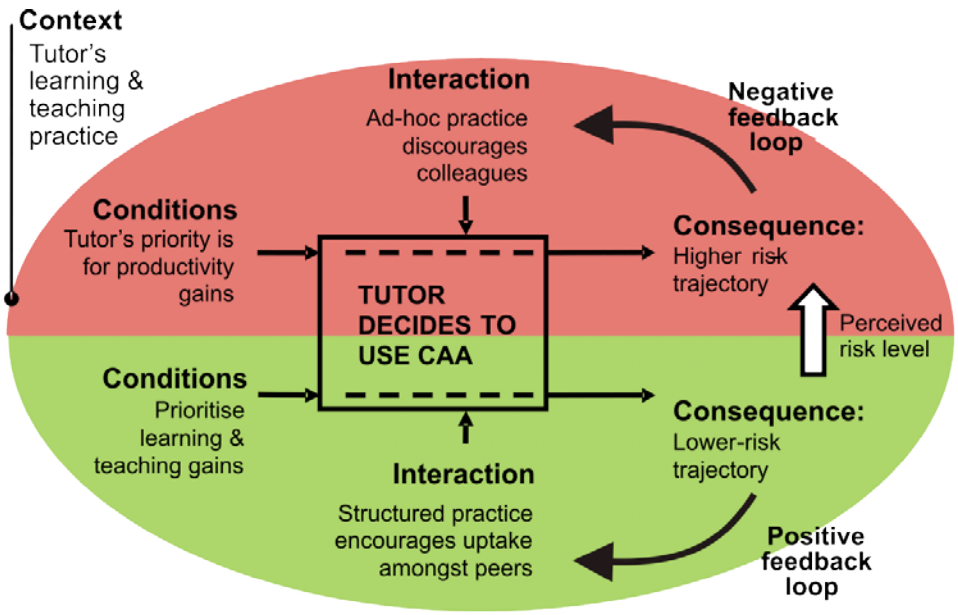

Figure 21: Core dual path theory

Slow burn?

Conversely, where tutors aimed primarily for pedagogical improvements they incurred much less risk and the resultant trajectories were characterised by routine use of small scale quizzes with an initial emphasis on low stakes testing such as formative and diagnostic applications. This sometimes progressed towards higher stakes testing on a larger scale:

> [Their] work on this... [began with] pilot development stages. We [proceeded] to running some Key and Basic Skills tests... with the prospect of widening this out to regular sessions, on-demand and then to high stakes options. (Learning technologist LtS6M002)

This staged, incremental approach was encouraged and supported by learning technologists who recognised the value for tutors of learning to use complex CAA tools in less critical applications. High stakes applications such as examinations were seen by learning technologists very much as risky undertakings which should be the final goal of CAA trajectories:

> [We] did a needs analysis of who wanted to use it... then my team... decided to pilot formative, summative...They wanted to go from there to using it for exams because they had already sort of built a little bit of confidence up. (Learning technologist LtN4F001)

Experienced tutors made similar points:

> For final summative assessment on-line, there are many technical difficulties to overcome and the whole operation is much more complex than a normal paper exam... [it's got to be] planned for. (Tutor AmO5F001)

Staged lower risk trajectories of this kind generally produced productivity gains and consequently diffusion tended to be steady rather than spectacular. Where tutors emulated this approach, they appeared to do so because they perceived a structured, methodical pattern of practice which would protect their investment in assessment materials and which might yield sustainable if modest productivity gains in the medium to long term:

> [We saw that] those that do them get instant feedback on their progress. (Tutor AmO3F001)

> [We saw] ...higher motivation of students through immediate feedback. Students persist until they get the right solution... (Tutor AMO5O002)

The reduced risk incurred by pedagogically-driven attitudes to CAA use and the accelerating effect this has on uptake by colleagues is shown in the lower half of the model (Figure 21).

Appendix J: Questionnaire used for interviews

Questions

What denotes ‘high quality’ in summative e-assessment?

What steps do you follow to create and use summative e-assessment?

How do you ensure: reliability, validity, security and accessibility of e-assessment?

How does the process of creating good quality e-assessment differ from the process of creating traditional assessment?

Please give us examples of good e-assessment; why these are ‘good’?

When you have heard of poor e-assessment, what has made it ‘poor’?

What feedback have you received from students who have taken e-assessments?

What advice would you give to others using summative e-assessment?

Do you have colleagues you feel could give us advice from the perspective of another role?

What research or other work has informed your thinking about summative e-assessment? And what further research would you like to see conducted?

Appendix K: Non-objective question design

Cliff Beevers

Building on an initial formative approach to e-assessment the scope and role of the computer in summative assessment has been explored through articles, book chapters and research projects like Pass-IT[25]. For over two decades e-assessment in numerate subjects has employed techniques of string evaluation or algebraic equivalence to test student answers in the form of mathematical expressions or equations. Often the questions posed contain random parameters which take on different values each time the question runs, requires a staged or stepped approach to question design and looks for mathematical expressions or equations as answers to the steps or parts of the question. This approach to e-assessment usually combines formative and summative roles with the same questions being employed in both modes — the only difference being the amount and level of feedback visible on the screen and the range allotted to the randoms. Moreover, this seam of work has built up a sound research foundation together with a growing reference list. One place for the interested reader to start would be the rich vein of articles developed from 2002 and lodged on the website of the UK Higher Education Mathematics Support Centre. These articles can be found online[26] or by visiting the website of the e-Assessment Association.[27]

An example will help to illustrate some of the issues in constructive non-objective question design:

Example: The circle C has diameter PQ with coordinates P(a, b) and Q(2c-a, 2d-b). Find the equation of the circle C.

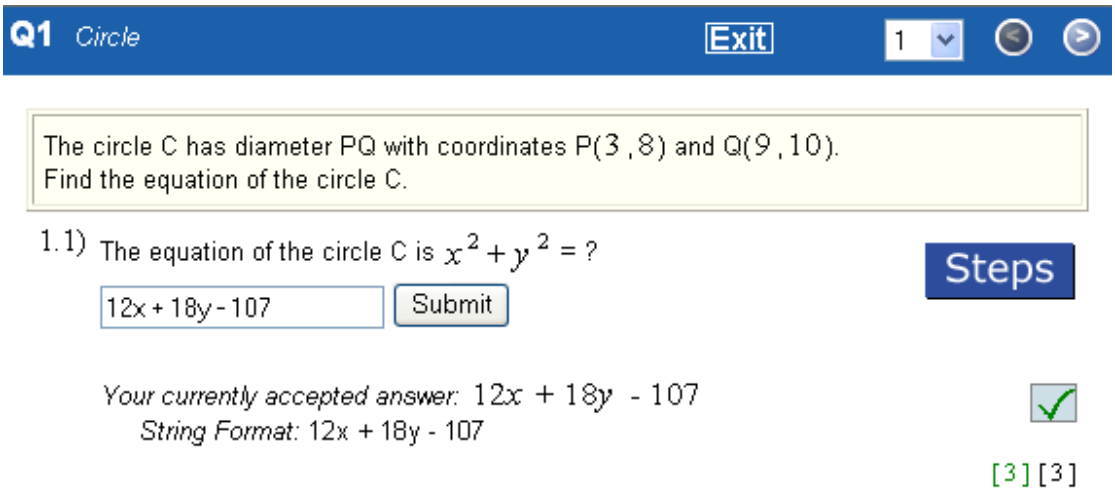

Figure 22: One instance of the circle question used in formative mode

[25] http://www.pass-it.org.uk/
[26] http://www.mathstore.ac.uk/index.php?pid=136
[27] http://www.e-assessment.com/

In this example the random parameters a, b, c and d can be given a range of values, say integers between -9 and 9 with the only restriction that a should not equal c, or b should not equal d (to ensure that P and Q are distinct points). Each time the question is displayed a student will receive different coordinates for P and Q giving in excess of 120,000 instances of the same question available to the question designer. Thus, this question provides a student with lots of formative practice. The example above requires knowledge, understanding and application to find the centre O and radius of the circle C and hence its equation. The student is then presented with two options: give the mathematical expression that answers the main part of the question shown at 1, or take a smaller step towards the answer.

Example: The circle C has diameter PQ with coordinates P(a, b) and Q(2c-a, 2d-b). Find the equation of the circle C.

1. The equation of the circle C is $x^2 + y^2 = ?\ 2cx + 2dy + a^2 + b^2 - 2ac - 2bd$

Take steps?

In this design a good student can answer this question as shown by typing in the equation of the circle and receive say 3 marks. However, the less able student might need to take smaller steps to reach the final answer. So, the example above with the Steps key pressed might reveal something like:

Example: The circle C has diameter PQ with coordinates P(a, b) and Q(2c-a, 2d-b). Find the equation of the circle C.

S1: The coordinates of the centre of the circle are ? (c, d)

S2: What is the radius of this circle ? $\sqrt{((a - c)^2 + (b - d)^2)}$

1. The equation of the circle C is $x^2 + y^2 = ?$

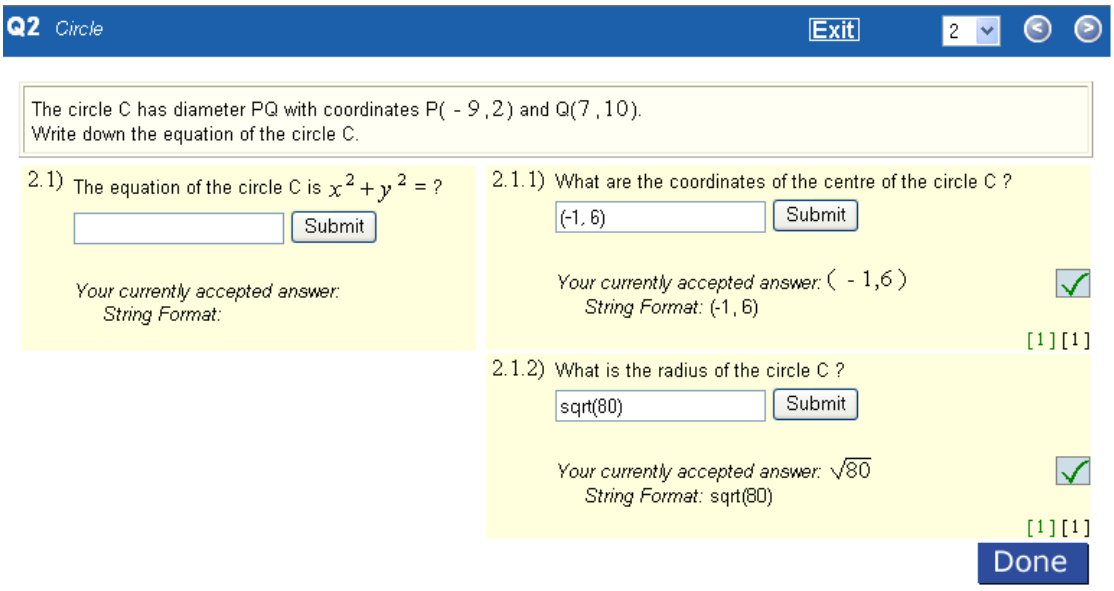

Figure 23: A second instance of the circle question showing formative mode with optional steps

The steps S1 and S2 might enable a student to respond in part to this question to obtain some partial credit. Indeed, the student can then return to the main question and provide the appropriate constructed response as an answer. This design enables

students to progress and score some partial credit. Moreover, with ticks and crosses on the screen in a formative mode, it can build up a student's confidence.

This design through optional steps lends itself well to the conversion of learning outcomes in a given topic. In the example a student might be expected to:

Determine the mid-point of the line PQ and hence find the centre of the circle (1 mark);

Know how to find the distance OP, the radius (1 mark); and

Construct the equation of a circle from the coordinates of its centre and the value of its radius (1 mark).

These three learning outcomes fit neatly with the two steps and key part design of this question. A fuller explanation of this technique of question design is well-documented in the article by Ashton and Youngson (2004). Variants of this question design have received scrutiny over the years in work by Beevers et al (1999) and McGuire et al (2002) in which issues of partial credit have been investigated.

There are comparisons of traditional and electronic versions of the same questions in the work of Greenwood et al (1998) and Sandene et al (2005) in secondary education in Northern Ireland and the United States respectively. The design outlined above provides a sound basis on which to research comparability between electronic and traditional versions of the same questions or whole examinations. Building on earlier work by Beevers et al (1995) and Fiddes et al (2002), a largescale study of comparability at a number of levels and subjects was undertaken in the Pass-IT project.[28] Major articles by Ashton et al (2005) and Ashton et al (2006) inform on further modifications to question design to ensure comparability with traditional examination measurements and hence retain the quality of summative e-assessment. One aspect of the research outcomes of Pass-IT can be demonstrated with a slight modification of the same circle example described above, Figure 23:

Example: The circle C has diameter PQ with coordinates P(a, b) and Q(2c-a, 2d-b). Find the equation of the circle C.

S0: The centre O of the circle lies halfway between P and Q.

S1: The coordinates of the centre of the circle are ? (c, d)

S2: What is the radius of this circle ? $\sqrt{((a-c)^2 + (b-d)^2)}$

1. The equation of the circle C is $x^2 + y^2 = ?$

[28] http://www.pass-it.org.uk/

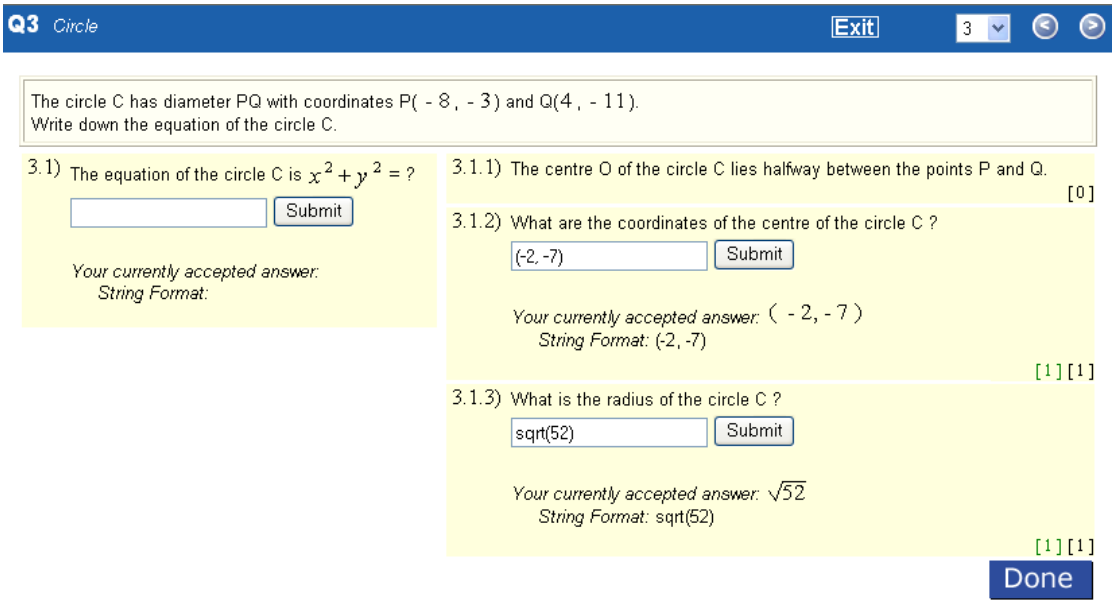

Figure 24: A third instance of the circle question in summative mode showing a student using steps.

This question, (Figure 24) has now been given an additional step S0 in which the strategy has been revealed on how to find the coordinates of the centre of the circle. So, it is recommended in summative mode that a part mark be forfeited on taking the steps. This modification of mark loss need not be applied in formative mode and indeed such a step as S0 can be classified as a hint which might well jog the student into completing this question. In its summative form however a part mark deduction led to a better comparison of traditional and electronic versions of the same question. Moreover, a more complex statistical theory has been proposed for these multi-staged questions and the interested reader is directed to the article by Korabinski et al (2008) for further details.

Another example of the approach described in this appendix has been active at Newcastle University for over four years in which e-assessment is available for all 1st year Mathematics modules. These are used for in-course assessment, 4 per module, contributing 10% of their module grade (see Foster (2004, 2007) for further details). It is now used for Mathematics, Statistics and Service Mathematics for all 1st and 2nd years honours modules at Newcastle University, see Fawcett et al (2008) for an extension to service teaching of statistics. One important part of this process at Newcastle was that e-Assessment in these modules became school policy. As part of this these assessments were moved from “belonging to the lecturer” to “belonging to the school”. There is both administrative and technical support with positive encouragement from the Head of Department. The original driver was cost saving on the marking of large numbers of text-based in course assessments by post-graduate students.

Moreover, where the assessment engine uses algebraic equivalence it is possible to test higher order skills in mathematical subjects. Finally, a recent article by Ashton (2008) contains additional examples of constructed response questions.

References

H S Ashton and M A Youngson, (2004). Creating questions for automatic assessment in mathematics. Retrieved February 2009 from http://ltsn.mathstore.ac.uk/articles/maths-caa-series/feb2004/index.shtml

H S Ashton, C E Beevers, A A Korabinski and M A Youngson (2006). Incorporating partial credit in computer-aided assessment of Mathematics in secondary education, *British Journal of Educational Technology*, 37(1), 93-119 (2006)

H S Ashton, C E Beevers, A A Korabinski and M A Youngson (2005). Investigating the Medium Effect in School Chemistry and College Computing National Examinations, *British Journal of Educational Technology*, 36(5), 771-787 (2005)

H S Ashton, (2008). Exploring e-Assessment: A practitioner's perspective. Retrieved February 2009 from http://www.elearning.ac.uk/features/e-assessment

C E Beevers, G R McGuire, G Stirling and D G Wild, (1995). Mathematical Ability Assessed by Computer, *J Comp Educ* 25, 123 - 132.

C E Beevers, M A Youngson, G R McGuire, D G Wild and D J Fiddes, (1999). Issues of Partial Credit in Mathematical Assessment by Computer, *Alt-J*, 7, p26 - 32

L Fawcett, W H Foster and A Youd, (2008). Using computer based assessments in a large statistics service course. *MSOR Connections*, 8(3), 45-48.

D J Fiddes, A A Korabinski, G R McGuire, M A Youngson and D McMillan (2002). Does the mode of delivery affect Mathematics examination results? *Alt-J* 10, 60-69 (2002)

W H Foster, (2004). Keeping control using computer aided assessment. Retrieved February 2009 from http://www.mathstore.ac.uk/articles/maths-caa-series/dec2004/index.shtml

W H Foster, (2007). Using computer based assessment in first year mathematics and statistics degree courses at Newcastle University. *MSOR Connections*, 7(3), 41-45

L Greenwood, U M Cole, F V McBride, H Morrison, P Cowan and M Lee, (1998). Can the same results be obtained using computer-mediated tests as for paper-based tests for National Curriculum assessment? *Int Conf Mathematics / Science Educ and Tech*, Vol. 2000, Issue. 1, 2000, pp 179-184.

A A Korabinski, M A Youngson and M McAlpine, (2008). Estimation of the accessibility of items and the confidence of candidates: A Rasch-based approach. *J. App. Measurement*, 9(1).

G R McGuire, M A Youngson, A A Korabinski and D McMillan (2002). Partial credit in mathematics exams - a comparison of traditional and CAA exams, *Proceedings 6th International CAA Conference*, Loughborough University (2002).

B Sandene, N Horkay, R E Bennett, N Allen, J Braswell, B Kaplan and A Oranje, (2005). Online Assessment in Mathematics and Writing: Reports From the NAEP Technology-Based Assessment Project, Research and Development Series, U.S. Department of Education, Institute of Education Sciences.