

**SOPHISTICATED TASKS IN E-
ASSESSMENT:
WHAT ARE THEY? AND WHAT ARE
THEIR BENEFITS?**

Andrew Boyle

Sophisticated Tasks in E-Assessment: What are they? And what are their benefits?

Andrew Boyle,
Research and Statistics team,
Qualifications and Curriculum Authority (QCA),
83 Piccadilly,
London W1J 8QA.
0207 509 5349
BoyleA@qca.org.uk
<http://www.qca.org.uk/>

Abstract

This paper asserts the importance of e-assessment. It further suggests that assessment questions and tasks will change substantially as the art of e-assessment progresses. The paper then exemplifies sophisticated e-assessment tasks, and seeks to identify aspects of a definition of them.

Next, some key claims for sophisticated e-assessment tasks are summarised and evaluated. These claims are:

- That sophisticated e-assessment tasks can be used to assess novel constructs
- That sophisticated e-assessment tasks can be used to address summative and formative assessment purposes.

In the final part of the paper, issues arising from the paper's findings are discussed and necessary areas for further research are noted.

The importance of e-assessment

E-assessment is important. It is an important aspect of several government initiatives. Government initiatives that concern e-assessment include the Department for Education and Skills (DfES)'s e-learning strategy. The objective of the e-learning strategy is '... to encourage a constructive national debate on how e-learning can contribute to the future role of education ...' (DfES, 2004). Assessment is an important part of the e-learning strategy. Also, the recently published 'Tomlinson report' emphasised the key role of e-assessment in the future of the education system (Working Group on 14 – 19 Reform, 2004, p. 63). The Government's White Paper on 14 – 19 education repeats Tomlinson's endorsement of e-assessment (DfES, 2005, p. 74).

E-assessment will be widely used in the near future. The Qualifications and Curriculum Authority (QCA) has produced a set of targets for Awarding Bodies

(QCA, 2004a) that, if successfully implemented, will make e-assessment commonplace in high-stakes assessment in the 14 – 19 sector of education by the end of the current decade. E-assessment (or Computer-Assisted Assessment) is also well established in the Higher Education sector (Sim et al, 2004).

The importance of sophisticated tasks in e-assessment

E-assessment will become an important and widely-used feature of the education system in the near future. Further, the types of questions and tasks used in near-future e-assessment will most likely be quite different from questions and tasks used in on-paper assessment, and in early implementations of computerised assessment.

This view is supported by the Tomlinson report, which states that:

E-assessment should not be construed as limited to quick multiple-choice testing; it has the potential to test learners in both structured and unstructured environments, with both short and long answer questions, as well as the ability to use techniques such as video clips to test a wider range of knowledge, skills and understanding than is possible at present. (Working Group on 14 – 19 Reform, 2004, p. 63; see also, DfES, 2005, p. 75).

Furthermore, a major project aimed at assessing all pupils in Information and Communications Technology (ICT) at the end of Key Stage 3 (the lower secondary phase of education in England) is premised on a model of assessment which requires pupils to demonstrate their ICT capabilities by performing tasks using a toolbox of bespoke applications and a virtual world of simulated web pages, documents, email messages and other assets (QCA, 2004b; Peppiat, 2004). The ways in which pupils work in the Key Stage 3 ICT test are very far removed from the ways in which they work in on-paper tests, or simple-item computerised tests.

These recent developments are inspired by a strand of research that has sought to predict the future of e-assessment. Some key propositions from this strand are outlined below.

Bunderson and colleagues suggest that, in 'intelligent measurement', more sophisticated assessment methods will become common. This will include the use of computerised versions of complex assessment techniques such as simulations and case studies (Bunderson et al, 1989, p. 369). Computerisation will allow these assessment methods to be carried out in a standardised and cost-effective manner (p. 370).

Bennett has written extensively about the future of e-assessment. For example, he conceives of a 'next generation' of assessment, in which the first thing to change will be the nature of test questions and tasks (Bennett, 1998, p. 5). For example, widespread use of high-quality computerised multimedia will permit radically different test tasks (see also: Bennett et al, 1999). In the final stage of development, Bennett envisages a 'reinvented' assessment landscape. In this environment, integrated learning and assessment tasks might include: intelligent tutors, microworlds and simulations (Bennett, 1998, p. 12). Content will be highly dynamic and adaptable to

take into account the particular interests, and strengths and weaknesses of individual learners (*ibid.*).

Other Bennett articles address different aspects of the future of e-assessment. For example, Bennett (2002) notes the practical drivers that have, so far, limited e-assessment to translations of pencil-and-paper tests. However, he believes that the 'emerging science of assessment design' will permit the use of simulations and other complex performance assessments. Such instruments should permit the awarding of summative assessment grades, and also output information that can be used to inform future learning (2002, p. 15).

There is support from other commentators for the contention that tasks in e-assessment will change substantially. Ridgway and McCusker (2004) suggest that computer presentation makes it easier to present new task types. Such novel tasks can display information dynamically, for example to model changes in situations over time. Pupils can also be presented with information in several formats, and be allowed to alternate between the formats. This flexibility of presentation means that pupils can interact realistically with data of considerable complexity (2004, p. 24).

Some literature cautions that sophisticated e-assessment might not be easily introduced into education systems with characteristics similar to those of the present day. For example, Raikes and Harding note that high-stakes e-examinations in England would have to operate in an environment with very variable ICT infrastructures. This would cause concerns that some pupils, in schools with weak ICT infrastructures, would be disadvantaged by innovative e-assessments. This would probably lead to test sponsors needing to provide a pencil-and-paper alternative to the e-assessment. Further, the requirement in English assessment that standards over time be maintained might militate against the rapid conversion of existing examinations to sophisticated e-formats (Raikes & Harding, 2003, p. 269).

It is important to acknowledge the significant inhibitors to the adoption of sophisticated-task e-assessment. However, the combination of significant current initiatives and credible expert predictions suggests that sophisticated tasks will have an important role in the near future of e-assessment.

Examples of sophisticated e-assessment tasks

There are many examples of sophisticated tasks in e-assessment. In this section three types of task are chosen to illustrate some facets of different types of sophisticated tasks.

Tripartite Interactive Assessment Development system (TRIADs)

The Tripartite Interactive Assessment Development system (TRIADs) has been developed by the Centre for Interactive Assessment Design (CIAD) at the University of Derby. TRIADs is widely used in United Kingdom Higher Education (HE) for the creation and delivery of e-assessments. The system permits test developers to choose from more than 30 question styles (Mackenzie, 1999, p. 1).

Mackenzie claims that his test delivery platform's ability to permit the creation of assessment tasks that contain multiple interaction styles is a major advantage. He shows the following screenshot of a task to support this claim (Mackenzie, 1999, p. 7):

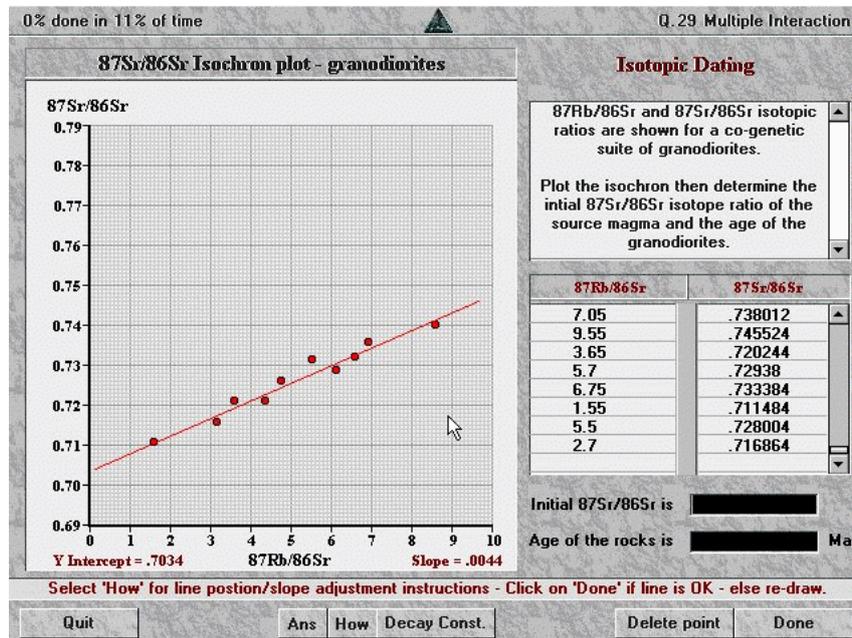


Figure 1. A sophisticated e-assessment task containing several interaction styles

In this task, students are required to plot points on the graph from data provided in the table. Then, they must draw by eye a line of best fit to join the plotted points. Next, they use the characteristics of the plotted line of best fit to derive a formula to calculate an isotope ratio (which they type into the box in the bottom right-hand corner of the screen).

World Class Tests

World Class Tests (WCTs) are optional assessments that can be taken by gifted and talented nine- and thirteen-year-olds in the United Kingdom, and other countries (World Class Arena, 2004; QCA, 2002). The sophisticated multimedia interface of WCT tasks is claimed to have several advantages:

- Simple animations can provide a more direct and less abstract representation of reality for pupils.
- Simulations can enable pupils to understand scientific process, such as the interaction between variables.
- The computer enables large amounts of data to be presented quickly.
- The computer can be used as a tool to explore complex data, or to model the effect of different responses to complex problems (QCA, 2004c).

Below is an example screen from a WCT task.

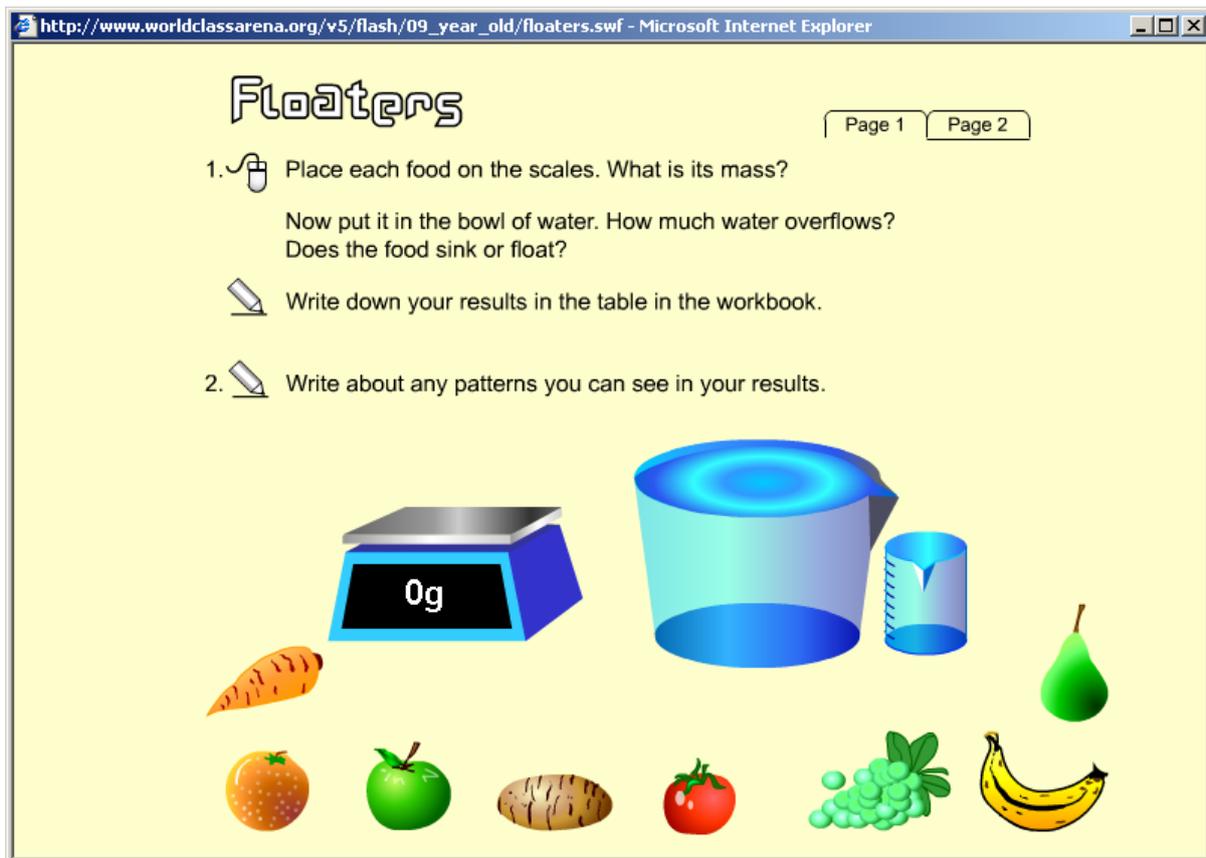


Figure 2. World Class Tests: A screen from the 'Floaters' task

The floaters task is highly visual, and will allow young students to be tested on an aspect of scientific method that would probably not be practicable for a large-scale pencil-and-paper test. It should be noted that this relatively ancient sophisticated task contains substantial elements in which students must express their answers on paper, and correspondingly there is a requirement for human marking.

Concept maps

In a comprehensive review of e-assessment task types for creative and critical thinking skills, Harlen and Deakin Crick found (amongst other things) that:

'Computer-based concept maps with automatic scoring can be used to provide summative assessment of critical and creative thinking and complex relationships.' Harlen and Deakin Crick (2003, p. 3)

Concept (or knowledge) mapping (Shavelsen et al, 1994; O'Neill & Klein, 1997; Araceli Ruiz-Primo et al, 1997) requires pupils to name concepts in a domain and to create links between named concepts. Variations on this assessment methodology can be made, for example by supplying pupils with the names of concepts, and/or information about the nature of the links between them (Osmundson et al, 1999). Concepts and links are represented graphically and can be clicked and dragged and located on the screen to provide a representation of knowledge in the area being assessed.

Taxonomies of sophisticated e-assessment tasks

The examples above give an initial illustration of what is meant by sophisticated e-assessment tasks. However, a small number of examples cannot define the characteristics of sophisticated tasks authoritatively. Some writers have created taxonomies of e-assessment tasks. Some exemplar taxonomies are outlined below.

Sim et al classify tasks into four groups depending upon the mode of interaction between a test taker and the computer. The four question styles that they discern are: point and click, move object, text entry and draw object (2004, p. 218).

Koch (1993) suggests four categories for sophisticated tasks:

- Traditional items with minor modifications
- Items with full use of graphics
- 'Multidimensional' tasks (i.e. tasks in which information has to be manipulated in several modes)
- Situated tasks (i.e. tasks with a high degree of 'real-world congruence').

Parshall et al. (2000, p. 130) propose a five-dimensional framework to categorise sophisticated e-assessment tasks (see also Zenisky & Sireci, 2002). Parshall and colleagues' five dimensions are:

- Item format
- Response action
- Media inclusion
- Level of interactivity
- Scoring algorithm

For Parshall and her colleagues, the category of item format basically pivots on whether the student has to select a response (as in a multiple-choice item) or construct a response (as in an essay question). Response action concerns the physical action that students have to carry out to interact with the computer; current tests might require a student to click a mouse, or to enter text through the keyboard. (See also Bennett et al, 2000, for examples of different response actions.) It is possible that, in future, more response actions will be available in tests – for example, with computers able to understand and score spoken responses (p. 135). Media inclusion refers to the extent to which e-assessments include still graphics, audio, video and animation. The 'level of interactivity' dimension refers to the extent of sophistication in interactions between students and the digital test content. Finally, scoring algorithm describes the range of ways in which sophisticated responses to tasks can be scored automatically.

Parshall et al suggest that, in practice, the five dimensions for describing sophisticated tasks are likely to be inter-related. For example, a task requiring complex interactions would require an informed approach to scoring (Parshall, et al, 2000, p. 130). Further, the five dimensions operate as continua; e-assessment tasks could be allocated on each continuum from simple to sophisticated, but it is not likely to be possible to define a 'cut point', before which all tasks are simple, and beyond which all are sophisticated.

Towards a definition of sophisticated e-assessment tasks

Sophisticated tasks are, necessarily, complex. Therefore, it is unlikely that a simple definition for them can be found (cf. Mackenzie, 2004, p. 1). However, it is possible to suggest some core features.

Firstly, sophisticated tasks should contain media-rich stimulus material (whether graphical, sound, video or animation). Secondly, the test taker should be required to interact with the stimulus material in a variety of ways. It follows that the work that a pupil produces in a sophisticated-task test will be more complex than work produced in a simple-item test (although not necessarily more complex than work produced in some pencil-and-paper task types – such as essays).

Several notes can be added to this list of core features. Firstly, for an e-assessment task to be considered sophisticated, it must contain several of the core features; a single feature on its own would not be enough to allow a task to be considered sophisticated. Secondly, the features of sophisticated tasks exist to distinguish them from their simple computerised or pencil-and-paper cousins. To some extent, the category of sophisticated tasks exists solely to emphasise the promise and novelty of the new medium. It is not likely that ‘sophistication’ can exist as a formally defined category, in isolation of discussion of the alleged merit of e-assessment tasks.

Possible benefits of sophisticated e-assessment tasks

Many commentators have alleged that sophisticated e-assessment tasks have considerable benefits over previously used methods. In the following section, two of the most important possible benefits of sophisticated e-assessment tasks are described and evaluated.

Measuring different things

Critics of tests made up of simple items have alleged that the abilities that are addressed by tests made up from simple items are based on a dated form of psychology. This form of psychology characterises a learner as a ‘collector of facts and skills, adding each to his repertoire more or less independently’ (Mislevy, 1993, p. 27). This view is at odds with more recent understanding of learning and experts’ cognitive activity, in which performance can be described as a complex interaction of heuristics, automaticity, metacognitive skills and schemas (*ibid.*).

Further, some writers have alleged that the implicit models within conventional tests do not represent the cognitive strategies that students use when taking any type of test. For example, success in a conventional test probably depends upon at least two dimensions: accuracy and speed (Snow & Lohman, 1993, p. 2). Also, analyses of tests that fail to account for the many and varied sources of difficulty in test questions either misrepresent students’ abilities (*ibid.* at p. 3, and see also Ahmed & Pollitt, 1999), or, at the very least, overlook potentially interesting information about abilities.

Finally, there is a curriculum coverage argument. If some ‘higher-order’ skills are not easily tested by simple item formats, but are valued in curriculums, then tests that

were based upon such items would not represent the curriculum adequately (Bennett, 1993, p. 100).

This latter argument would be particularly important if it could be shown that many modern curriculums were beginning to emphasise new types of skills, knowledges and abilities. Indeed, some have argued that such a change can be observed in several educational jurisdictions. Ridgway and McCusker (2003) detect a 'shared rhetoric on education' in several countries. This shared vision includes:

- An increased emphasis on mathematics and science
- The increased use of ICT
- An increased emphasis on 'higher-order thinking' skills, such as problem solving and communication skills (2003, p. 310).

Ridgway and McCusker go on to argue that there is a complex relationship between technology and the development of new cognitive abilities (2003, p. 312). But they find it reasonable that novel technologies require people to employ novel cognitive skills.

Harlen and Deakin Crick consider that critical and creative thinking are 'key aspects of higher level thinking' (2003, p. 52). Further, they cite authorities who claim that conventional testing does not provide adequate assessment of these important skills (2003, p. 53) and that sophisticated multimedia presentations do have the potential (for example, because of their non-linear and dynamic representations of information) to provide the necessary scaffolding and support to permit meaningful assessment of 'higher-order skills' (2003, p. 53).

Several other projects using sophisticated e-assessment tasks have asserted the medium's suitability for assessing 'novel skills and abilities'. For example, the new Key Stage 3 ICT test attempts to assess ICT capability, which is defined as follows:

'ICT capability is about having the technical and cognitive proficiency to access, use and communicate information using technological tools.

Learners demonstrate this capability by purposefully applying technology to solve problems, analyse information, develop ideas, create models and exchange information.

They are discriminating in their use of information and ICT tools.' (Brant, personal communication; see also: Peppiat, 2004)

This construct is contrasted with ICT skills, which are the technical competences necessary to do simple tasks using commonly-used software applications. Several tests using conventional items address ICT skills.

Maughan and Mackenzie report on an assessment that uses an on-screen simulation of a microscope that purports to assess process skills as well as the success of outcomes in Biology (2004, p. 206). Similarly, Rayne and Baggott advocate the use of 'modern assessment authoring tools – which may incorporate simulations, animations, game scenarios, etc.' (2004, p. 307) to assess procedural knowledge (how to do something) and the conceptual knowledge underpinning

procedure. Finally, Duke-Williams and King (2004) report on the use of a range of 'drag-and-drop' type questions to assess higher-level learning outcomes in Bloom's taxonomy (although see McCormick (2004, p. 126) for a criticism of the uninformed use of Bloom's taxonomy).

Thus, this section has set out a range of criticisms of the underlying – and often implicit – models of cognition that underpin conventional-item tests. These criticisms seem intuitively valid. However, insofar as sophisticated tasks in e-assessment depart from outdated psychological models, then thought has to be given to the new mental models that underpin novel task types. This vast area will present significant challenges to researchers. It will be necessary to account for the cognitive implications of sophisticated tasks when specifying the assessment model (Pellegrino et al, 2001, p. 179), and when validating tests (Araceli Ruiz-Primo et al, 2004).

It will be important to avoid construct-irrelevant variance; that is, to make sure that sophistication and complexity in task user interfaces do not amount to 'nuisance variables', and thus detract from validity (Huff & Sireci, 2001, p. 19). Further, it will be crucial to evaluate the differential impact of new task styles for learners with varying backgrounds and dispositions.

The rapid spread of e-technologies in recent years also suggests that now is a good time to re-explore the objects of assessment. An in-depth analysis of the full range of educational goals that can be addressed by sophisticated e-assessment is beyond the scope of this short paper. However, it seems likely that some goals of sophisticated e-assessment will be more sustainable than others. For example, the desire to assess problem solving, and critical and creative thinking skills seems to be widespread. In contrast, some proposed assessment goals may require more thinking. For example, it may well be desirable to assess 'process skills'. But, in major substantive domains the process of solving problems, or creating sophisticated work products, is complex; there are many ways to solve a problem or to produce a piece of work. A great danger would be if a test favoured a particular path through a problem over other legitimate solutions. If this were to be the case, and a prescriptive summative assessment model were to be in place, then an imperfect complex e-assessment of process skills would run the risk of either disadvantaging students with certain learning styles, and/or skewing classroom activity substantially.

Addressing different assessment purposes

A substantial recent strand of assessment research in England has concerned the various purposes to which assessments may be put. Researchers have described the benefits that can accrue for pupils' learning when teachers employ formative assessment effectively (Black & Wiliam, 1998). Work has also been done with classroom teachers to ascertain and disseminate good practice in aspects of the process of formative assessment (Black & Wiliam, 2003). For example, principles for providing feedback that will nurture students' learning have been put forward (Nicol & MacFarlane-Dick, 2004). Finally, researchers have investigated the effects of summative assessment on students' motivation (Deakin & Harlen Crick, 2002).

The e-assessment literature has addressed similar issues. Writers have seen the potential of e-assessment to provide useful feedback on aspects of sophisticated tasks (Baggott & Rayne, 2002; Ridgway et al, 2004). Well-designed formative information is felt to be useful for students and to be more supportive of their learning than the right|wrong information that can be obtained from a simple item (Maughan et al, 2001). Moreover, the range of question styles and multimedia presentations that are available in sophisticated tasks have been argued to provide the possibility of feedback in a range of visual styles and modes of interaction, these likely to be more consistent with students' styles of learning (Mackenzie, 1999).

However, two reservations to this optimism must be expressed. The first is economic, the second concerns the professional skills of teachers. These reservations are described in turn. E-assessment is expensive to develop. Also, the costs of e-assessment tend to be shifted forward in the assessment cycle. Thus, whilst e-assessment can potentially save money in activities carried out after the administration of a test (such as the transportation or marking of scripts), large budgets seem – at the time of writing – to be inevitable for the test development phase (cf. Ricketts et al, 2003). This is especially true for sophisticated-task e-assessment. Large assessment organisations are overwhelmingly the most likely institutions to have the available resource to commit to such projects. Such institutions are generally focussed on summative assessment, and often have a conservative, risk-averse orientation (Wainer, 2000). If such institutions remain the best equipped to develop sophisticated-task e-assessments, then it bodes ill for the method's use for innovative formative assessment.

The second reservation is a mirror of the first. Writing good test items is difficult. Research in Higher Education, where e-assessment is widely used by classroom practitioners who are not assessment specialists, suggests that the skills of good item writing do not necessarily co-occur with the skills of good teaching (Boyle et al, 2002; Boyle & O'Hare, 2003). Some have claimed that the difficulty for non-specialist practitioners in writing test tasks is particularly pronounced when it is necessary to create complex tasks (McLaughlin et al, 2004)¹. It is certainly true that high-quality test items take a long time to develop, and need to go through several iterations of review and editing (Sainsbury, 2002).

Thus, at the present time, sophisticated tasks for e-assessment may be characterised as expensive and slow to develop, and not easily written by a non-specialist teacher. If these characterisations are true, and for the period of time that they remain true, then the widely-shared aspiration for sophisticated e-assessment to provide high-quality formative assessment might not be realised.

Thus, sophisticated-task e-assessment may not be a technology that teachers can easily take ownership of. This means that information from sophisticated e-assessment tests may – necessarily – continue to come from remote central sources. If this is so, it will stand in opposition to one of the essences of high-quality

¹ The counterclaim is that sophisticated tasks resemble teaching and learning materials more closely than conventional test items do. As such, it should be easier for teachers to write materials for sophisticated e-assessments (O'Hare, personal communication). This argument requires one to accept that teachers are, or will be in a foreseeable future, skilled developers of multimedia learning materials.

formative assessment; that assessment practices belong to teachers and learners and are closely integrated with classroom practice.

Discussion

This paper has asserted the importance of e-assessment; in particular, it has stated that sophisticated e-assessment tasks will become widely used in the near future. Subsequently, the paper has exemplified and sought to define sophisticated tasks. Finally, it has gone on to address two of the key claimed benefits of sophisticated tasks in e-assessment; that is, that they can assess new skills, and that they can be used for formative assessment.

The paper has not found a clear-cut definition of sophisticated tasks; rather, it is suggested that the term 'sophisticated tasks' might serve only as a device to distinguish some kinds of tasks from the purportedly reactionary use of simple multiple-choice questions in early computerised testing implementations. Similarly, whilst the supposed benefits of sophisticated e-assessment tasks for measuring new constructs, and for facilitating formative assessment are rehearsed, critiques of the alleged benefits are launched in both cases.

The findings of this paper might suggest, then, that one should not accept the key claims for sophisticated e-assessment tasks. In fact, this is not the case. Scepticism should be exercised when evaluating the potential benefits of any educational innovation, but the central assertion that sophisticated tasks will soon be ubiquitous, and that they will bring real benefits, is accepted. However, many of the papers on sophisticated tasks to date have asserted these benefits, rather than analysing the necessary concepts, or grappling with practical problems that will need to be thought through in order for successful implementation to be achieved. This is perhaps to be expected, as such early papers needed to alert a readership to the potential of sophisticated tasks. The challenge now is to move on to informed implementation.

Two main benefits have been asserted for sophisticated e-assessment tasks: firstly, that they will allow the assessment of 'higher-order skills' more effectively than traditional methods, and, secondly, that they will facilitate formative assessment. These two arguments are often confounded in the literature, when in fact they are distinct. Further, it is felt that the former is the stronger argument for sophisticated tasks. The limitations of conventional test items for assessing high-order skills are well documented. Empirical evidence of sophisticated tasks' ability to do this more successfully is, as yet, sparse, but there are several projects currently underway that are realistically optimistic that they will soon be able to provide such evidence.

In contrast, the many assertions of sophisticated-task e-assessment's ability to facilitate good formative assessment seem less secure. This is particularly because of the complexity of authoring complex tasks, and the corresponding likelihood that such tasks will continue to be created by large, central organisations. This likely outcome flies in the face of the requirement that formative assessment be low-tech and easy to use in everyday classrooms. It may be that other, more mundane ICT applications are suitable for formative assessment – for example using tracking software in word processors to provide useful feedback.

This short paper has not been able to address everything of relevance to sophisticated e-assessment tasks. The successful use of these tasks will require researchers to consider a wide range of issues, such as the methods by which scores are derived for each task, the measurement models that can be used to summarise scores on tasks to provide meaningful information about performances on tests, and issues in the validation of tests. These issues will be crucial in future research into this area.

Acknowledgements

Tim Oates: Head of QCA Research & Statistics team

References

Ahmed, A. and Pollitt, A. (1999) *Curriculum demands and question difficulty*. Paper presented at IAEA Conference, Slovenia, May 1999. <<http://www.ucles.org.uk/assessmentdirector/articles/confproceedingsetc/IAEA1999AAAP>> (24 February 2005).

Araceli Ruiz-Primo, M., Shavelson, R. and Elise Shultz, S. (1997) *On the validity of concept map-based assessment interpretations: an experiment testing the assumption of hierarchical concept maps in science*, CSE Technical Report 455, 1997 <http://www.cse.ucla.edu/products/reports_set.htm> (24 February 2005).

Araceli Ruiz-Primo, M., Li, M., Yin, Y., Shavelson, R.J., Vanides, J., Schultz, S. and Ayala, C. (2004) *Concept maps as an assessment tool: a framework for examining their cognitive validity*. Paper presented at Paper Presented at the American Educational Research Association (AERA) Annual Meeting, April 2004, San Diego, CA. <<http://www.stanford.edu/dept/SUSE/SEAL/AERA%202004/>> (24 February 2005).

Baggott, G. and Rayne, R. (2002) *The online assessment and feedback project: project rationale* <<http://www.bbk.ac.uk/olaaf/extras/rationale-extra.pdf>> (24 February 2005).

Bennett, R.E. (1993) Towards intelligent assessment: an integration of constructed-response testing, artificial intelligence and model-based measurement in Frederiksen, N., Mislevy, R.J. and Bejar, I.I. (eds) *Test theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bennett, R.E. (1998) *Reinventing Assessment: Speculations on the Future of Large-Scale Educational Testing*. Princeton, NJ: Educational Testing Service Policy Information Center <<ftp://ftp.ets.org/pub/res/reinvent.pdf>> (24 February 2005).

Bennett, R.E. (2002) *Inexorable and Inevitable: The Continuing Story of Technology and Assessment*. The Journal of Technology, Learning and Assessment (JTLA) 1 (1) <<http://www.bc.edu/research/intasc/jtla/journal/v1n1.shtml>> (24 February 2005).

Bennett, R.E., Goodman, M., Hessinger, J., Liggett, J., Marshall, G., Kahn, H., and Zack, J. (1999) *Using multimedia in large-scale computer-based testing programs* Computers in Human Behavior **15** 283-294.

Bennett, R.E., Morley, M. and Quardt, D. (2000) *Three response types for broadening the conception of mathematical problem solving in computerised tests*, Applied Psychological Measurement **24** (4) 294 – 309.

Black, P. and Wiliam, D. (1998) *Assessment and classroom learning*. Assessment in Education **5** (1) 7 – 73.

Black, P. and Wiliam, D. (2003) *In praise of educational research: formative assessment*. British Educational Research Journal **29** (5) 623 – 637.

Boyle, A. and O'Hare, D. (2003). *Finding appropriate methods to assure quality Computer-Based Assessment development in UK Higher Education* in Christie, J. (ed) Seventh International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 5th and 6th July 2003 <<http://www.caaconference.com>> (24 February 2005).

Boyle, A., Hutchison, D., O'Hare, D. and Patterson, A. (2002) *Item selection and application in higher education* in Danson, M. (ed) Sixth International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 9th and 10th July 2002 <www.lboro.ac.uk/service/ltd/flicaa/conf2002/pdfs/ohare_d1.pdf> (24 February 2005).

Bunderson, C.V., Inouye, D.K. and Olsen, J.B. (1989). The four generations of computerized educational measurement in Linn, R.L. (ed) *Educational Measurement*. Third edn. New York, NY: American Council on Education.

Department for Education and Skills (DfES) (2004) *e-learning strategy*. <<http://www.dfes.gov.uk/elearningstrategy/index.cfm>> (24 February 2005).

Department for Education and Skills (DfES) (2005) *White Paper on 14-19 Education and Skills*. <<http://www.dfes.gov.uk/publications/14-19educationandskills/>> (24 February 2005).

Duke-Williams, E. and King, T. (2001) *Using computer-aided assessment to test higher level learning outcomes* in Danson, M. and Eabry, C. (eds) Fifth International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 2nd and 3rd July 2001 <<http://www.caaconference.com>> (24 February 2005).

Harlen, W. and Deakin Crick, R. (2002) *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. <http://eppi.ioe.ac.uk/EPPIWeb/dl.aspx?Path=reel%5Creview_groups%5Cassessment%5Cass_rv1&FileName=ass_rv1> (24 February 2005).

Harlen W. and Deakin Crick, R. (2003) *A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills*

<http://eppi.ioe.ac.uk/EPPIWeb/dl.aspx?Path=reel%5Creview_groups%5Cassessme nt%5Cass_rv2&FileName=ass_rv2> (24 February 2005).

Huff, K.L. and Sireci, S.G. (2001) *Validity issues in computer based testing*. *Educational Measurement: Issues and Practice* **20** (3) 16–25.

Koch, D.A. (1993) *Testing goes graphical*. *Journal of Interactive Instruction Development* **5** 14 – 21.

Mackenzie, D. (1999) *Recent developments in the Tripartite Interactive Assessment Delivery system (TRIADs)*. Paper presented at the third International Computer-Assisted Assessment Conference, July 1999. <<http://www.derby.ac.uk/ciad/lough99pr.html>> (24 February 2005).

Mackenzie, D. (2004) *Online assessment: quality production and delivery for Higher Education*. <<http://www.enhancementthemes.ac.uk/uploads/documents/Mackenziepaperrevised.doc>> (24 February 2005).

Maughan, S. and Mackenzie, D. (2004) *Bioscope: The assessment of process and outcomes using the TRIAD system* in Ashby, M. (ed.) Eighth International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 6th and 7th July 2004 <<http://www.caaconference.com/>> (24 February 2005).

Maughan, S., Peet, D. and Willmott, A. (2001) *On-line formative assessment item banking and learning support* in Danson, M. and Eabry, C. (eds) Fifth International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 2nd and 3rd July 2001 <<http://www.caaconference.com>> (24 February 2005).

McCormick, R. (2004) *ICT and pupil assessment*. *The Curriculum Journal* **15** (2) 115 – 137.

McLaughlin, P.J., Fowell, S.L., Dangerfield, P.H., Newton, D.J. and Perry, S.E. (2004) Development of computerised assessment (TRIADs) in an undergraduate medical school in O'Hare, D. and Mackenzie, D. (eds) *Advances in Computer Aided Assessment*. SEDA: Birmingham.

Mislevy, R.J. (1993) Foundations of a new test theory in Frederiksen, N., Mislevy, R.J. and Bejar, I.I. (eds) *Test theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nicol, D. and Macfarlane-Dick, D. (2004) Rethinking formative assessment in HE: a theoretical model and seven principles of good feedback practice in Juwah, C., MacFarlane-Dick, D., Matthew, B. Nicol, D., Ross, D. and Smith, B. *Enhancing student learning through effective formative feedback*. The Higher Education Academy: York. <http://www.heacademy.ac.uk/resources.asp?id=353&process=full_record§ion=generic> (24 February 2005).

O'Neil, H. and Klein, D.C.D. (1997) *Feasibility of Machine Scoring of Concept Maps*, CSE Technical Report 460. <http://www.cse.ucla.edu/products/reports_set.htm> (24 February 2005).

Osmundson, E., Chung, G., Herl, H.E. and Klein, D.C.D. (1999) *Knowledge Mapping in the Classroom: A Tool for Examining the Development of Students' Conceptual Understandings*, CSE Technical Report 507. <http://www.cse.ucla.edu/products/reports_set.htm> (24 February 2005).

Parshall, C.G., Davey, T. and Pashley, P.J. (2000) Innovative item types for computerized testing in Van der Linden, W.J. and Glas, C.A.W. (eds) *Computerized Adaptive Testing: Theory and practice*. Dordrecht, NL: Kluwer Academic Publishers.

Pellegrino, J.W., Chudowsky, N. and Glaser, R. (eds) (2001) *Knowing What Students Know: the Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Peppiat, M. (2004) *Data-rich reporting and standards setting following rules-based marking* in Ashby, M. (ed) Eighth International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 6th and 7th July 2004 <<http://www.caaconference.com/>> (24 February 2005).

Qualifications and Curriculum Authority (QCA) (2002) *Assessing Gifted and Talented Children: papers presented at an international seminar, London, February, 2001*. London: QCA.

Qualifications and Curriculum Authority (QCA) (2004a) *QCA's e-assessment vision* <<http://www.qca.org.uk/7192.html>> (24 February 2005).

Qualifications and Curriculum Authority (QCA) (2004b) *On-screen test for ICT at Key Stage 3* <<http://www.qca.org.uk/7280.html>> (24 February 2005).

Qualifications and Curriculum Authority (QCA) (2004c) *World Class Arena* <<http://www.qca.org.uk/7282.html>> (24 February 2005).

Raikes, N. and Harding, R. (2003) *The Horseless Carriage Stage: replacing conventional measures*. *Assessment in Education* **10** (3) 267 – 277.

Rayne, R. and Baggott, G. (2004) *Computer-based and computer-assisted tests to assess procedural and conceptual knowledge* in Ashby, M. (ed) Eighth International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 6th and 7th July 2004 <<http://www.caaconference.com/>> (24 February 2005).

Ricketts, C., Filmore, P., Lowry, R. and Wilks, S. (2003) *How should we measure the costs of computer-aided assessment?* in Christie, J. (ed) Seventh International Computer Assisted Assessment (CAA) Conference Proceedings, Loughborough University, 5th and 6th July 2003 <<http://www.caaconference.com>> (24 February 2005).

Ridgway, J. and McCusker, S. (2003) *Using computers to assess new educational goals*, *Assessment in Education: Principles, Policy and Practice* **10** (3) 309 – 328.

Ridgway, J., McCusker, S. and Pead, D. (2004) *Literature review of e-assessment*. <http://www.nestafuturelab.org/research/reviews/10_01.htm> (24 February 2005).

Sainsbury, M. (2002) *Test development research: a short guide*. TOPIC **25** item 9.

Shavelson, R., Lang, H. and Lewin, B. (1994) *On concept maps as potential "authentic" assessments in science*, CSE Technical Report 388, <http://www.cse.ucla.edu/products/reports_set.htm> (24 February 2005).

Sim, G., Holifield, P. and Brown, M. (2004) *Implementation of computer assisted assessment: lessons from the literature*. *Association for Learning Technology-Journal (ALT-J)*, **12** (3) 215 – 229.

Snow, R.E. and Lohman, D.F. (1993) *Cognitive psychology, new test design and new test theory: an introduction* in Frederiksen, N., Mislevy, R.J. and Bejar, I.I. (eds) *Test theory for a new generation of tests*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wainer, H. (2000) *CATs: whither and whence*. *Psicológica* **21** 121 – 133. <<http://www.uv.es/psicologica/paraARCHIVES/2000.html>> (24 February 2005).

Working Group on 14 – 19 Reform (2004) *14 – 19 Curriculum and Qualifications Reform: Final Report of the Working Group on 14 – 19 Reform*. Annesley, Notts.: DfES.

World Class Arena (2004) *World Class Arena: Challenging minds, recognising talent* <<http://www.worldclassarena.org/v5/default.htm>> (24 February 2005).

Zenisky, A.L. and Sireci, S.G. (2002) *Technological innovations in large-scale assessment*. *Applied Measurement in Education* **15** (4) 337 – 362.