E-assessment and innovation
John Winkley

This report was produced by AlphaPlus for Becta (Year of Publication) and is Crown Copyright, 2010. This material has been developed from publications produced by Becta or its predecessor organisations which was made available under the Open Government Licence. (http://www.nationalarchives.gov.uk/doc/open-government-licence/open-government-licence.htm)
About the author

John Winkley is an experienced e-learning and e-assessment business manager with a track record of business growth in educational technology. He has specialised in developing innovative systems for education and assessment, working with a range of public and private sector clients through the four companies he has helped to build: BTL Group Ltd, The Virtual College plc, 3Square Solutions Ltd, and AlphaPlus Consultancy Ltd.

This work has involved developing and deploying innovative solutions for e-assessment in schools, and the further education and lifelong learning sectors. Projects range from large scale high-stakes examinations systems and content to highly innovative approaches to formative assessment. He has spoken widely at UK and international conferences on e-assessment.

Email: John.winkley@alphaplusconsultancy.co.uk
Contents

Introduction ........................................................................................................................................... 4
On-screen examinations ......................................................................................................................... 4
Beyond multiple-choice questions – innovative items ................................................................. 5
Use of audio ........................................................................................................................................ 6
Text editing ........................................................................................................................................ 7
Adaptive assessment .......................................................................................................................... 8
Simulations and serious gaming ........................................................................................................ 9
Innovation in assessment marking .................................................................................................... 12
  Short text answer marking ............................................................................................................. 12
  Essay marking ............................................................................................................................... 14
Spoken language assessment marking ............................................................................................ 14
Foreign language translation marking ............................................................................................... 15
Mathematics marking ....................................................................................................................... 16
Assessment for Learning and e-assessment ..................................................................................... 16
  The AsTTle Project, New Zealand ................................................................................................. 17
E-marking ........................................................................................................................................... 18
Use of mobile devices ....................................................................................................................... 18
The key benefits of e-assessment as used today .......................................................................... 19
The future for e-assessment ............................................................................................................. 20
References and further reading .......................................................................................................... 22
Introduction

The ways in which we live, learn and work have been transformed by technology over the last few decades and it is therefore not surprising that people in the 21st century need new skills to cope in the information society. Alongside these new skills are new ways of assessing them.

The term ‘e-assessment’ began to be used widely from the late 1990s. Ofqual\(^1\) defines e-assessment as:

…the use of electronic systems for the development, operation and delivery of accredited qualification assessment or the collection of performance evidence, which contributes to the awarding of a unit or an accredited qualification.

In looking at the types of e-assessment that institutions use, it is important to consider summative, ‘high stakes’ assessments\(^2\) – examinations where the results matter to the candidate, the institution (for example, school league tables) and/or society, and hence are both carefully regulated and of public interest. In contrast, assessments for learning (as they are called in school, or ‘formative assessment’ more widely) – those embedded within learning activities, are designed solely for learners and teachers to monitor and plan their work. While the technologies employed in these two settings may be similar, approaches to risk and innovation are very different.

On-screen examinations

On-screen examinations – summative high-stakes e-assessment – are very much the public face of e-assessment and have been around for many years: IT engineers and aircraft pilots have been taking their professional qualifications with organisations such as Microsoft, Cisco and the Federal Aviation Administration for around 20 years. However, it is only over the last ten years or so that the use of on-screen examinations has become relatively common in UK public sector education, and today it remains more the exception than the norm as far as the majority of examinations are concerned.

The Driving Standards Agency offers around 1.6 million on-screen tests a year to learner car drivers and motorcyclists. The Home Office offers the ‘Life in the UK tests’ on screen to applicants for British citizenship. Many professional qualifications, such as those in accountancy, medicine and law, require candidates to sit on-screen assessments.

\(^1\) http://www.ofqual.gov.uk/64.aspx
\(^2\) http://en.wikipedia.org/wiki/High-stakes_assessment
Within the public education system in the UK, there is wider use of e-assessment within post-compulsory provision than in schools. Around one million candidates take on-screen assessments each year in Skills for Life and Key Skills. Newly qualifying teachers have to pass on-screen tests of their numeracy, literacy and ICT skills before they can qualify to teach in schools. Almost all colleges have experience of offering on-screen testing regularly to students. For example, in a survey in 2008 of 20 FE colleges chosen at random, all had been using e-assessment for examinations for three years or more, with over 80 per cent offering more than 1000 e-tests a year. Within higher education, many HE institutions use on-screen tests to assess course units, particularly in the first year of a degree programme.

In 2004, QCA set out a blueprint for e-assessment: proposed objectives to achieve by 2009, that ‘all new qualifications would include an option for on-screen assessment’ and that ‘most GCSE and A-level exams would be available on screen’. These targets have not been reached: while many providers use e-assessment, its use is very limited within GCSE and A-levels. So, e-assessment is widely used in vocational qualifications for testing knowledge and understanding (City and Guilds recently announced their 3,000,000th on-screen test). Its use in academic qualifications is growing, but is still at the level of pilot work in most cases.

**Beyond multiple-choice questions – innovative items**

It’s true that most e-assessment today uses computer-marked MCQs (multiple-choice questions), yet there is also a rich community of e-assessment development work offering more sophisticated questions and marking which takes e-assessment well beyond the simple, objective MCQ.

There are numerous types of on-screen questions which retain the benefits of computer-marking, but avoid the tightly closed nature of multiple-choice questions. The advantages over paper-based equivalents (other than the electronic marking) relate either to interactivity – providing the candidate with powerful and realistic ways to interact with the question resources and present their response, or to rich media – providing a variety of realistic and stimulating content to allow more authentic assessment activity. The benefits of these items in terms of assessment include improved motivation and engagement from students, and improved face and content validity. A number of examples are included below.

---

3 Becta E-Assessment Readiness Survey in FE, 2008, unpublished  
6 Face Validity is a subjective measure of whether the test ‘appears valid’ to candidates and the general public. Often, this is an important feature in test acceptance/public buy-in. Content Validity is a measure of how well a test covers the topic. It’s common for tests not to test some sections of the curriculum, either due to time constraints or because they are hard to test in an exam (group-work, for
Interactive Microscope as part of GCE Examinations

Professor Don Mackenzie of I4L worked with the team at Cambridge Assessment to develop a fully functioning virtual microscope to support A and AS level exams. The assessment captures student interactions with the microscope, allowing the examiner to cross-reference with diagrams the student sketches during the exam, ensuring the diagrams have not simply been memorised.

Use of audio
The question on the right, taken from the 2010 Skills for Life Literacy Survey, is very simple from a technical perspective, but provides a very effective way of testing listening and spelling skills which is simply not possible with a paper assessment. This and similar simple innovations can broaden the extent to which the assessment covers the curriculum quite considerably.

example) but if too much is left out then the test starts to lose its overall validity – its effectiveness at measuring what it is supposed to.

7 http://www.i4learn.co.uk/content.asp?menuId=4
The Driving Hazard Perception Test

As part of the Car and Motorcycle Theory Tests, candidates have to watch a video of everyday road scenes, and identify potential and developing hazards. In the example here, the cyclist has previously been identified as a potential hazard, and is now a developing hazard as he is looking over his shoulder and beginning to pull out. The test is about road scanning – identifying potential hazards early, and tracking them to see if they develop. Candidates click on the screen at both points, and score depending on how quickly they spot the hazard.

Text editing

In the examples here, of literacy and numeracy questions, candidates use interfaces that allow constructed responses but with a degree of restriction on the options for entry which allow computer marking to be possible.
None of these examples use particularly advanced technology. The three key operational challenges in producing such material are (1) ensuring the technology is used only to enhance the assessment and not as gimmicks or irrelevant visual improvement, (2) providing powerful content tools which allow such items to be generated cost effectively without large costs for customisation, and (3) ensuring that question content is re-useable for cost-effectiveness as e-assessment content is significantly more expensive than traditional content to produce.

**Adaptive assessment**

As a candidate progresses through an adaptive assessment, the computer chooses which questions to ask next, based on the candidate’s responses to previous questions. Broadly speaking, if the candidate is doing well, then harder questions are presented. Adaptive assessments have three main advantages:

- Within a given space of time, they can measure a student’s capabilities more accurately than a traditional linear test, because more of the questions are targeted close to the student’s ability, so greater differentiation is available. This not only speeds up testing, it can also provide personalised topic-level profiles to support personalised learning planning.
- Adaptive assessments can be particularly useful where the candidature has a very wide range of ability, for example when assessing learners for course placement. Linear tests inevitably tend to pitch questions at the ‘average’ level, which reduces the differentiation at the extremes.
- Because a candidate taking an adaptive assessment is generally attempting questions at or around their level, the tests tend to be more motivating, so completion rates and effort levels are higher.

There are a number of technical methods of implementing adaptive assessment. One example is shown on the next page (taken from the LSIS\(^8\) initial assessment tool for Skills for Life Literacy which is used to ensure students are placed on the right level of course). Other methods commonly used include dynamic item selection from a pool of items\(^9\) based on statistical methods (IRT\(^10\) generally). Finally, it is worth noting that there are disadvantages with adaptive testing: a large number of items are required and these items have to be well understood in terms of their difficulty to ensure that the adaptivity functions correctly.

The example of an adaptive assessment (below) is used for LSIS’ Initial Assessment of Literacy Skills. A typical example of its use is placing students on the correct course on arrival at an FE college.

---

\(^8\) LSIS – The Learning and Skills Improvement Service [http://www.lsis.org.uk]

\(^9\) http://en.wikipedia.org/wiki/Computerized_adaptive_testing

\(^10\) http://en.wikipedia.org/wiki/Item_response_theory
The range of skills assessed within this single test is broad (from writing a simple sentence at Entry Level 1 through to Level 2 which is equivalent to GCSE Grade C). Students start with Question 1 in Segment 1. The diagram shows candidate routes through literacy assessment, showing branching points for candidates achieving lower and higher scores. The ‘snakes and ladders’ structure is used to route learners upwards, downwards or straight on according to their performance at the end of each cluster of questions, enabling learners to climb to the highest level even after a poor start.

Simulations and serious gaming

The use of ICT-based simulations in education is well known and understood – simulations of practical systems are particularly useful where the student’s interaction with the real thing is impractical, for reasons of danger (such as in medicine, in flying a plane, or operating a nuclear reactor), too time consuming (as in slow chemical reactions or economic systems), or too expensive. For example, the Scottish Qualifications Authority provides a simulation of the work environment for students in full-time education as part of its Skills for Work programme. In many cases simulators can provide diagnostic and/or summative assessment about the student’s performance as a by-product of the learning experience (an attractive option in many cases, reducing the apparent burden of assessment). In this section, we look at a number of examples taken from different subject domains and levels, and using different types of simulation technology.

‘Serious gaming’ refers to games with an educational purpose. However, ‘normal’ video games have a learning objective as well: teaching the player how to play the game. Many also have rigorous pass/fail mechanisms too. However, particular attention is placed in serious games to measuring educational outcomes, so rigorous assessment is a key area of development in serious gaming.

http://www.sqa.org.uk/sqa/36537.2262.html
Primum – medical diagnosis
Primum is used as one of three parts of the United States Medical Licensing Examination (USMLE), which is designed to test would-be doctors’ ability to treat patients in a practical setting, competencies that were previously examined at the bedside by supervision and observation. It is a high-stakes, computer-based case study simulation where candidates are presented with authentic problems and are asked to treat a simulated patient on screen.

Candidates, using free text entry (candidates are free to make any diagnosis or investigation – they do not pick from a list), receive information, conduct examinations and order tests and treatments, to which the electronic patient will respond. A candidate’s performance is assessed against model responses using a regression-based, automated scoring procedure.

EcoMUVE
The EcoMUVE\textsuperscript{12} project at the Harvard Graduate School of Education provides a Multi-User Virtual Environment (MUVE) for work on ecosystems in middle school science. It allows multiple students to access virtual worlds simultaneously, interacting with other students and with computer-based agents to facilitate collaborative learning activities of various types. A typical scenario would be ‘Students discover a fish kill at the pond and have to visit the pond (virtually, in the MUVE) at different points in time to figure out what happened’. The events have complex causality and require a combination of collaboration, research, logic and reasoning, analytic and reporting skills. Partial computer marking is supported by the use of tokenised XML.

\textsuperscript{12} \url{http://www.ecomuve.org/index.html}
**FloodSim**

One example of serious gaming is FloodSim\(^{13}\) (illustrated right) which ‘...puts you in control of all flood policy decisions and spending in the UK for 3 years. Whether it’s deciding how much money to allocate to flood defences, deciding where to build houses, or how best to inform people about the risk of flooding, you are in control. [It] aims to raise awareness of the vast number of issues surrounding flood policy and Government expenditure and to increase citizen engagement through an accessible simulation.’

Rather like adaptive assessment, the assessment within simulations tends to be fully embedded as a feature of the simulation. This of course helps students take ownership of their learning and is a key part of the game-style motivation to ‘try again’. However, it can also become ‘too implicit’ so that learners don’t necessarily think carefully about the detail of the outcome they’ve achieved, so serious games can often have support resources to assist in making the experience more ‘formative’.

Simulation technology is quite varied. Some use generic simulation engines (such as Expert Systems\(^{14}\)), which are then tailored for the domain and for educational purpose (in contrast to, for example, supporting expert decision-

---

\(^{13}\) [http://floodsim.com/](http://floodsim.com/), developed by PlayGen.

making). Other simulations are computationally relatively simple, relying rather on rich and varied media (the World Class Tests and Geography are examples of this) and can be hand crafted using relatively simple software.

In general, simulations can be costly to develop, particularly if they are to appear authentic – with realistic resources and settings, and with sufficient sophistication to respond convincingly to a wide range of student/candidate input. Work in the future is likely to focus on developing simulation engines for assessment which allow different scenarios to be produced cost effectively for particular domains.

Innovation in assessment marking

Aside from enriching the student’s experience during the assessment, there is also considerable innovative work in the field of computer marking, mostly in the area of free text, that is, marking students’ typed responses in the form of phrases, sentences, paragraphs and essays. The opportunity for both immediate feedback to students and cost-saving on human marking makes this a very attractive area for investigation, although it is technically very challenging, and current implementations can lead students towards writing essays with a particular structure which gets high marks from the system. The requirement for heavy processing power for linguistic analysis makes marking on the desktop computer impractical, but projects are increasingly using web services to submit candidate responses for marking at a central, powerful system.

There are a number of subtly different approaches taken in this area – the section below describes some of the main examples.

Short text answer marking

Short text answers\textsuperscript{15} are answers that have to be constructed rather than selected, ranging from phrases up to (rarely) three to four sentences. A typical example, from GCSE science, would be ‘Describe and explain what happens to the mass of a candle when it burns’. They are widely used in UK academic assessment at Key Stage, GCSE and GCE testing, and are appealing, compared to multiple choice, because they require a student to construct rather than pick an answer and allow for partial marks.

\textsuperscript{15} See the section on ‘Automatic scoring of constructed, short text responses’ in the report on the Review of Advanced e-Assessment Techniques project [http://www.jisc.ac.uk/publications/documents/raeatfinalreport.aspx].
Engines exist to mark these questions, but two issues have to be considered:

- They work best with questions producing convergent answers, i.e. where there is a limited (though large) set of correct answers.
- They are not appropriate where knowledge and understanding is being tested, and where content is important rather than style.

Systems generally use technologies such as spellcheckers, thesauri, and latent semantic analysis to match the student's response to template answers (coping with differences in tense and grammar, synonyms etc.), and as a result require both considerable effort in setting up the template answers and training the system on a range of trial responses. However, noting these limitations, the quality of marking can be very high, and can also provide partial marks and helpful feedback to students in formative settings.

Such systems are currently limited to responses of one or occasionally two, sentences. It is tempting to think that as technology progresses, the boundaries will extend to cover computer marking of longer answers, for example, paragraph responses. However, the problem is that the range of possible answers expands very quickly with increasing quantity of text (the widely observed combinatorial explosion problem).

Some systems, which provide a 'confidence level' identifying how certain the computer is that it has understood the response and marked accurately, allow a mixed economy of computer marking (for the run-of-the-mill responses) and human marking (for the unusual ones) and offer interesting possibilities which some awarding organisations are exploring.

There are a number of e-assessment products which score short-text responses. A recent JISC review\(^\text{16}\) of advanced e-assessment focused on two systems, Openmark\(^\text{17}\) and Intelligent Assessment Technologies\(^\text{18}\) as used at the Open University Centre for Open Learning of Mathematics\(^\text{19}\), Science, Computing and Technology, and Dundee University\(^\text{20}\).

\(^{16}\) Five case studies of advanced e-assessment have been researched. Each was selected and organised around a specific e-assessment theme, as follows: formative assessment, higher order skills, combining human and computer marking, automatic scoring of foreign languages, and automatic scoring of free text. [http://www.jisc.ac.uk/media/documents/projects/raeat_finalreport.pdf]

\(^{17}\) http://www.open.ac.uk/openmarkexamples/

\(^{18}\) http://www.intelligentassessment.com/

\(^{19}\) http://www.open.ac.uk/colmsct/

\(^{20}\) http://www.dundee.ac.uk/
Essay marking

The testing of ability to construct a sustained argument, which is often tested in essay questions, is not well addressed through short text assessment. The best-known program for marking essays is ETS\textsuperscript{21} e-rater (as used on the GMAT\textsuperscript{22} examination) which compares correlations of essay quality (style, vocabulary and length, for example) with those of a battery of pre-graded scripts, finds the best match and scores accordingly. It is admirably suited to a tradition where scripts are routinely doubly marked by examiners but is not used as the sole marking system. There is no sense in which the program ‘understands’ the content of the essay.

Spoken language assessment marking

There are many settings where students’ ability to speak a particular language needs to be assessed. The Versant\textsuperscript{23} system is used for assessing spoken English as part of citizenship tests, employment and professional certification testing, etc.

The tests are delivered over the telephone or the web. During the test, the examinee hears a number of prompts and questions designed to elicit responses in the target language. The questions are designed carefully to limit the range of possible correct responses. For example, the candidate might see three sentences written on the test paper:

1. World Air 891, request descent.
2. World Air 891, maintain flight level 280 expect descent shortly.

and will hear “Please read sentence number three.” Other examples include requiring the candidate to correct spoken errors. For example, the examinee might hear the following two speakers:

(Speaker 1) “Charlie Romeo 4013, continue descent to flight level 110, report passing 150.”

(Speaker 2) “Descending to flight level 10 thousand, report passing 15 thousand, Charlie Romeo 4013.”

and be expected to correct the errors in the second speaker’s response, using accepted air traffic controller phraseology.

\textsuperscript{21}\url{http://www.ets.org}, See also a paper on e-rater in \textit{JTLA} \[\url{http://escholarship.bc.edu/jtla/vol4/3/}\]
\textsuperscript{22}\url{http://en.wikipedia.org/wiki/Graduate_Management_Admission_Test}
\textsuperscript{23}\url{http://www.versanttest.co.uk}
Responses are scored by comparing the examinee’s response to a database of proficient responses using digital signal analysis and speech recognition. This enables the system to respond ‘intelligently’ to the dialect of the test taker, without requiring test takers to use any form of standard pronunciation.

**Foreign language translation marking**

The University of Kent provides a formative e-assessment system which marks students’ translation of sentences from English to other modern languages. It uses an automatic scoring approach to provide the student taking the test with a marked-up error analysis of their first attempt at translation, then permits the student to submit a second response.

In the example below, a candidate has been asked to translate “The stubborn lawyer thanks the brave nurse” into French. A typical response and mark-up is shown below.

**Question 25**

**Question:**

- The stubborn lawyer (fem) thanks the brave nurse (fem)

**Attempt 1 (score 25%):**

- L’ouvrière tetu_ <se> remercie [___] courageux infirmière

**Amend your answer here:**

L’ouvrière tetu_se remercie courageux infirmière

Submit answer

This mark-up uses a classification system to help students understand the nature of their errors, but does not provide hints or guidance as to the correct translation. The system identifies the following categories of error:

- Incorrect words are shown in **bold red** font
- Missing letters and erroneous spelling/conjugation in otherwise correct words are shown in **bold pink** font
- Words which are superfluous are shown in `<green angle brackets>`

---

24 See the section on LISC [http://www.jisc.ac.uk/publications/documents/raefinalreport.aspx]
The system uses string-based matching for the marking, although Natural Language Processing is a common alternative for systems of this type. It is slower to set up lessons for LISC than for parser-based systems (where no correct answers are required), but the benefit is that LISC copes much better with poor input than parser-based systems (which need a way to predict errors and can't handle unforeseen combinations). As such the system requires more upfront investment than traditional approaches, which is a common feature of e-assessment innovation, and, coupled with user scepticism about benefits, can often lead to patchy and slow adoption.

Mathematics marking

A number of systems which provide marking of mathematical process, enabling students to present their working and then marking it in such a way as to award method marks and cope with alternative solutions, or equivalent mathematical forms.

Assessment for Learning and e-assessment

E-assessment for learning is an area where pedagogy is evolving rapidly alongside technology. It is also a mode of assessment which is situated at the heart of classroom teaching and learning. These two factors introduce an element of risk to adoption as changes to the core teaching and learning activities are necessary. Nevertheless, formative e-assessment also holds the promise of significant improvement in student achievement.  

A number of technology systems have sought to address these issues, by providing tailored adaptive assessments for groups of students, and producing actionable results for both students and teachers (focusing on organising the students into groups for customised teaching, for example). Two examples are worthy of mention:

- The AsTTle project in New Zealand (Maths, English and Maori)
- The Maths for Schools project in Northern Ireland.

---

In both cases, adaptive on-screen tests\(^{28}\) are prepared using IRT\(^{29}\) statistical methods. Students’ responses are analysed to produce individual spiky profiles of strengths and weaknesses for students, and then grouped into clusters of students with similar performances to provide practical teaching and class organisation recommendations for teachers. In the case of AsTTle, as results are uploaded to a central server, the teacher can also benchmark the class against school groups nationally.

**The AsTTle Project, New Zealand**

AsTTle enables teachers to create and analyse tests for literacy and numeracy (in English and Maori) presented onscreen or on paper. The resulting reports show what students know, what gaps they have in their learning, and what they need to learn next.

The screenshot shows the profile of a cohort by topic and attainment, placing students into four groups for each topic: those with gaps (red), below the achievement threshold (blue), at or just above the achievement threshold (green), and strong (yellow).

In the example above it is clear that (1) a large number of students are struggling with vocabulary, and (2) understanding is much weaker than the lower level skills.

---


The second image is a dashboard showing local against national performance in curriculum areas.

**E-marking**

Over the last five years or so, the awarding bodies offering GCSE and A levels have been introducing ICT to the processes of marking and standardising exam scripts. Whereas in the past, candidates’ scripts were collected and posted from school to the awarding body, and then on to the examiners’ homes, the scripts are now scanned, and distributed electronically for the examiners to mark on screen.

Although the technology for on-screen marking is not used widely for on-screen tests at the moment, it is clear that students’ typed answers to on-screen questions could be marked using this system. Some awarding bodies are investigating this as a way of offering on-demand assessments with rapid turn-around of results for human-marked questions for Functional Skills – bringing together the benefits of near-immediate feedback and the marking of longer answers and process skills.

**Electronic Voting Systems**

Following the theme of formative assessment – the use of assessment to take action to improve learning – the use of EVS systems has increased considerably, particularly in HE. A typical scenario is that students in a lecture are provided with handheld clickers such as the Activote pictured on the right to respond to questions asked by the lecturer. In addition to opportunities for interactivity, if the questions are well designed, they provide the lecturer with immediate real-time feedback on

---

whether the teaching is ‘sinking in’ and, with wrong answers, what the commonly held misconceptions are.

It also provides students with information about how they’re doing without exposing their scores to the rest of the class, as well as stimulating peer learning and support after the lecture. Effective use requires lecturer training – writing good diagnostic assessments and knowing how and when to use them isn’t easy – but there are now many examples of voting systems transforming the taught element of degree programmes.

**Use of mobile devices**

Allowing assessment to be fully embedded in the learning experience is particularly important where the learning is field-based rather than in the classroom. For example, the use of mobile devices allows students and tutors to record responses using photographs and spoken audio – much quicker than writing – while they are doing their fieldwork and then review it later. Examples of projects in this area include e-scape^32^ and the ALPS project^33^. Projects such as these blur the boundaries between e-portfolios and e-assessment, but, for e-assessment purposes, their key benefit is that they allow evidence to be captured with minimal interruption to the learning activity which then provides a rich basis for a formative or summative assessment activity.

**The key benefits of e-assessment as used today**

It’s worth reviewing some of the main benefits,^34^ as they point the way to where emerging technologies might best be applied:

**Benefit 1: immediate feedback**
Where tests are marked by computer, the results are available immediately, and can provide information not only about the overall result, but details of strong and weak areas.

**Benefit 2: improved assessment validity**
Assessment validity is a measure of ‘fitness for purpose’ or ‘authenticity’ – the extent to which the assessment activity is a fair measure of the actual skill or understanding. On-screen assessments can be more valid when they contain rich information (not just text, but sound and video, for example) and allow the candidate to interact with the information and present their information as flexibly as they wish.

---


^33^ [http://www.alps-cetl.ac.uk/suite.html](http://www.alps-cetl.ac.uk/suite.html)

^34^ It is worth noting that (1) not all these benefits apply to all e-assessment use, and (2) poorly designed or implemented e-assessment may achieve no benefits at all – the gains to be had are not ‘automatic’.
Benefit 3: increased flexibility
Assessments can be provided at a greater range of locations and times.

Benefit 4: more efficient and environmentally friendly administration
E-assessment, and the associated e-administration of candidates, reduces the administration burden on centres.

Benefit 5: candidates generally like it
Findings are overwhelmingly that where e-assessment is implemented well (with due account taken of personal preferences and the access requirements of those with physical or learning disabilities), the student response is positive.

The future for e-assessment

There are currently four key areas where e-assessment has potential to be transformational:

1 – developing the benefits of immediate results and improved feedback
While computer-based assessment has limitations compared to a skilled teacher’s careful assessment of a student, the practicalities of teacher–student ratios and the availability of immediate targeted feedback make formative e-assessment a very powerful tool for practical personalisation and learner empowerment.

2 – the benefits of the interactivity and richness of ICT for more authentic assessment
As the power of ICT increases, new options emerge for students to interact with question material and present their answers, and for computers to undertake aspects of the marking of these sophisticated responses (thereby enabling the benefits in theme 1 to be achieved). Such innovations push the boundaries of technology and pedagogy and inevitably result in assessments which have no paper-test equivalent.

3 – e-assessment as an enabler of better assessment
Quality of assessment remains a big political and educational issue. The availability of immediate and more detailed information about assessments allows and requires assessing institutions to be more proactive in managing their assessments.

4 – e-assessment for all
The fact is that most e-assessments today are produced not by the teacher but by third parties – awarding organisations and publishers. In the future, teachers will use tools to create and customise their own assessments, designed with their own specific purpose and audience in mind, thereby ensuring that assessment is fully embedded as a formative resource in the curriculum. When e-assessment first started, it was seen as a low-grade means of automating yes/no and similar question types. Now, it has a major role in all areas of education – placement, diagnostics, assessment for learning, diversity and inclusiveness, pedagogy, summative
assessment and awards, quality assurance, and personalisation – bringing immediacy, thoroughness, reliability and validity. In future, as all sections of the educational world learn from the more dynamic pace set by higher education, there will be a major shift in its use, from being a rarity to becoming the standard for assessment at all stages from primary to postgraduate. Its prevalence and relatively low cost will help to drive up standards and speed the responsiveness of educational institutions to changes in technology, knowledge and society.
References and further reading

AlphaPlus e-assessment white paper
[http://www.alphaplusconsultancy.co.uk/images/stories/PDFs/alphaplus_eassessment.pdf]

The Assessment and Teaching of 21st Century Skills project
[http://www.atc21s.org/home/]

Futurelab (2004 and 2007), E-assessment – an update on research, policy and practice
[http://www.futurelab.org.uk/resources/publications-reports-articles/literature-reviews/Literature-Review204]

JISC guide to effective practice with e-assessment
http://www.jisc.ac.uk/media/documents/themes/elearning/effpraceassess.pdf

JISC (2009), Advanced E-Assessment Techniques – a study
[http://www.jisc.ac.uk/publications/documents/raeatfinalreport.aspx]

JISC and QCA (2006), The e-assessment glossary
[http://www.jisc.ac.uk/media/documents/themes/elearning/eassess_glossary_extend edv101.pdf]

Ofqual (2009), Principles and Practice of On-demand Testing

Ofqual, e-futures website [http://e-assessment.org.uk/]

Acknowledgements

The author would like to thank the following organisations and individuals for providing resources for this article:

Andy McAnulla, BTL Group Ltd

Alison Fowler, The University of Kent

Ellen Arbuthnott, The Driving Standards Agency

Julia Landon, Promethean Ltd

Kam Star, PlayGen Ltd

The asTTle Project