

# Review of Advanced e-Assessment Techniques (RAeAT) Final Report



WORLD CLASS ARENA LIMITED

### **Project Team**

Martin Ripley  
Robert Harding  
Hakan Redif  
Jim Ridgway  
Jeremy Tafler

Contact: Martin Ripley  
[Martin@worldclassarena.net](mailto:Martin@worldclassarena.net)

Date: March 2009

3 Hampstead West  
224 Iverson Road  
London  
NW6 2HX

+44 (0) 20 7624 1418  
+44 (0) 20 7624 1615  
[info@worldclassarena.net](mailto:info@worldclassarena.net)

[www.worldclassarena.net](http://www.worldclassarena.net)

# Review of Advanced e-Assessment Techniques (RAeAT) Final Report

## Table of Contents

Acknowledgements	page 2
Executive Summary	page 2
Background	page 4
Aims and objectives	page 4
Methodology and Implementation	page 5
Output and results	page 7
Outcomes and Impact	page 11
Conclusions and Recommendations	page 13
Implications for the future	page 14
References	page 15
Appendixes	page 16

## Acknowledgements

This project and subsequent report was part of the JISC e-learning programme co-funded by the JISC Capital Programme and by Martin Ripley.

The Project Team would like to acknowledge and thank the following individuals/institutions who were our case studies and/or provided valuable input and expertise into the case studies:

John Hattie and Debra Masters at Visible Learning Labs, New Zealand

Dr Mary McGee at the University of Manchester

Brian Clauser at the National Board of Medical Examiners, USA

Alison Fowler at the University of Kent

Jared Bernstein and Alastair Van Moere at Pearson Versant

Sally Jordon and Phil Butcher at the Open University

Walter Williamson, John McEwen, Tom Mitchell at the University of Dundee

Andrew Thravesat at GLAssessment

Jerry Podmore and Laura Lassman at Doublestruck

We would also wish to thank Myles Danson and John Winkley for their invaluable guidance and review of our efforts.

## Executive Summary

There is exciting and effective practice in e-assessment in HEIs in the UK but take-up is patchy. This is due to a number of well-rehearsed factors but there are signs that progress being made in the development of e-assessment packages may encourage the take-up of these techniques more widely.

It is apparent from this study that e-assessment programmes have needed champions, both in the sense of the developer (often university-based, usually working in spare hours stolen from family or leisure time) and in the sense of institutional indulgence. Without these champions, potential academic end-users have found it unappealing to try and scale the walls of the new technology.

What is changing is usability. Where previously much of the preparatory work had to be done by third party or other technically expert staff, programs are increasingly providing end-user academics with the tools to implement their own e-assessment. The projects described in this report and five companion case studies show that benefits accrue. High quality, valid assessment can be performed which would not be economic without e-assessment; deadlines can be reached which would be otherwise unattainable. Not only is marking workload decreased, but information flow is increased. Tutors can easily check the efficacy of their teaching programmes; students can receive instant feedback on their learning; plagiarism is detected and discouraged; the impact of curriculum changes can be gauged.

Change is attractive when it meets problems that are already appreciated. One of the obstacles to wider acceptance of e-assessment is that it addresses issues that it has been all too easy to ignore. But increased demands for transparency, for timely feedback to students and for measuring teaching quality have raised the stakes. E-assessment offers the means of establishing students' misconceptions and of correcting them, of demonstrating and improving teacher effectiveness.

The aim of this project is to provide JISC and its community with a review of state-of-the-art techniques in e-assessment which should be considered for application in UK Higher Education (HE) and Further Education (FE) settings. This review will contribute to other programme work to define and develop a roadmap of e-assessment research and development activities. The objectives of the study are:

- To develop a picture of the breadth of advanced e-assessment techniques in use around the world, and undertake a technical and educational evaluation of a selection of the most promising;
- To present short case studies describing the technical, educational and operational details of a selection of the most significant examples;
- To look at wider generic technologies and investigate how and when these might be applied to e-assessment;
- To draw conclusions about how the most relevant of these could be applied to UK FE and HE settings, and provide advice to JISC and its community about such application could be developed.

The Review of Advanced e-Assessment Techniques project began by considering what constituted an advanced technique. "Advanced" refers to techniques that are used in isolated or restricted domains, and which have successfully applied technology to create an assessment tool. "Advanced" does not necessarily imply "newness".

The first stage within the project was to collate top-level descriptive information about potential candidates for further study in the project. This long-list ultimately contained over ninety

projects. It was a surprise to the project team how few previously unknown advanced e-assessment projects came to light through the trawls for information. The community of experts and departments using e-assessment is small, and this continues to have implications for scaling e-assessment and for stimulating the growth of additional innovative approaches.

The project team had expected to find secure examples of e-assessment techniques being developed within Web2 environments. Although environments such as Second Life are in use as social and learning environments, this project found little evidence of summative or formative assessment applications. We did identify a number of peer-to-peer assessment tools (such as Caspar at Bournemouth University and CAP at Glamorgan University), but considered that the generalisable issues relating to peer-to-peer have been well documented through the WebPA project. We also collected information about several other projects – such as TRIADS and eSCAPE – which certainly fulfil our criteria for “advanced techniques”, but we did not include further study of these as much has already been published. ePortfolios generally were not included as this assessment approach has been reviewed by a number of other JISC projects.

During the initial collection of information about advanced e-assessment techniques, the project team sought a range of projects covering as many different aspects of the assessment process as possible. We divided assessment business processes into three broad stages:

Pre-hoc process	This stage covers all processes prior to students completing a test, including processes such as authoring test items and preparing/printing the test papers.
Test administration	This stage covers all processes involved in students sitting a test or examination on a specific date.
Post-hoc processes	This stage covers all business processes that take place after the student has completed the test and might include, for example, marking and moderation.

In order to cover as broad a range as possible of these processes, we investigated a number of international projects, and looked at projects from the schools and professional training sectors.

A catalogue of all the Advanced e-Assessment Projects has been published at <http://www.dur.ac.uk/smart.centre1/aeac>

A total of five case studies have also been researched and written. Each of the case studies was selected and organised either around a specific e-assessment theme, or to investigate and describe a single e-assessment project. The case studies were as follows:

## Formative assessment

The focus of this case study is AsTTle. The main function of AsTTle is to provide users (tutors or teachers) with a system which enables them to specify and select test items to configure a test. The selection might be predicated on a focus on a particular curriculum domain, for example, or on the demand of the test items. The benefits of such a system derive from the ability to provide teachers with access to large, high-quality and calibrated item banks whilst also providing them with the flexibility to create test forms which closely reflect their teaching. Interestingly, AsTTle is an application developed in HE (in New Zealand), but providing services to school teachers.

<http://www.tki.org.nz/r/asttle/>

## Higher Order Skills

Some advocates of e-assessment point to the potential of computers to support simulation and scenario-based assessment. There are few examples of this category of e-assessment being developed successfully, especially not in high stakes testing contexts. Primum is an exception. It provides an assessment of trainee medical practitioners' ability in making medical diagnoses when presented with a fictitious patient exhibiting a number of symptoms. This automated assessment has been designed to provide an authentic and reliable assessment at a price that compares favourably with the alternative – human scored evaluation at patients' bedsides.

<http://www.usmle.org/Orientation/2009/menu.html>

## Combining human and computer marking

Few e-assessment systems are designed to drive efficient marking by automatically sorting and classifying students' responses, then directing the sorted listing of responses to human markers. Assessment21 does. The benefit of such an approach is that it delivers more efficient marking (a human marker can browse sorted responses, applying one judgement to multiple, identical responses) as well as supporting greater consistency in marking (a marker views all similarly constructed responses at one time and applies a single, common judgement to all of those similar responses).

[www.assessment21.com](http://www.assessment21.com)

## Automatic scoring of foreign language textual and spoken response

Computers have only limited ability to analyse and score the grammatical complexity of written language, and accuracy and fluency in the spoken word. Two e-assessment applications which do are LISC and Versant. LISC provides a detailed analysis of the grammatical accuracy of sentences translated into a second language by students, supporting students in their understanding and application of grammatical rules. Versant provides an automated spoken language proficiency assessment

with applications as diverse as Dutch government citizenship tests, university entrance fluency tests, and air-traffic controllers' proficiency tests. By designing assessments of highly focused skill domains, both of these e-assessments deliver effective assessments of skills that are prohibitively expensive to provide in a human-scored system.

<http://www.cs.kent.ac.uk/people/staff/amlf/call/html/menu.html>

<http://www.ordinate.com>

<http://htk.eng.cam.ac.uk/>

## Automatic scoring of constructed, short text responses.

There are a number of e-assessment products which score short-text responses. This case study provides a survey of current practice, with a particular focus on two systems – Automark and IAT. The strengths and limitations of short-test scoring are well understood and, although not a panacea, short-text scoring is a technically strong area of e-assessment. Short-text scoring systems provide the benefits of using constructed responses items (which can provide valid assessments of students' understanding and application of knowledge) and provide the technology-driven benefits of efficiency and accuracy.

<http://www.open.ac.uk/colmsct/>

[www.dundee.ac.uk](http://www.dundee.ac.uk)

This review of advanced e-assessment techniques has found examples of innovative and well researched e-assessment techniques. However, there is little evidence that e-assessment is having a substantial and positive impact on learning and teaching in UK universities. The examples of advanced e-assessment techniques studies in this project suggest that e-assessment can provide some solutions to concerns regarding assessment practice in higher education in the UK:

- e-assessment can offer opportunities to broaden the range of assessment methods used in universities and colleges, where currently the monoculture of essays continues to prevail;
- e-assessment offers opportunities to derive efficiencies in assessment – by automating the scoring of some types of item; by providing better diagnostic feedback; and by providing more flexibility over the timing of assessment to students and their tutors.
- e-assessment approaches could also provide routes to delivering greater consistency in assessment methods and standards between institutions.

It is beyond the scope of this report to speculate why advanced e-assessment techniques are not more prevalent in HE. The aim has been to find examples of e-assessment from which JISC and individual institutions can gain confidence about the education value and technical robustness of e-assessment systems

## Background

The aim of the wider JISC e-Learning programme is to enable UK further and higher education institutions to create better learning environments. Its vision is of a world where learners, teachers, researchers and wider institutional stakeholders use technology to enhance the overall educational experience by improving flexibility and creativity and by encouraging comprehensive and diverse personal, high quality learning, teaching and research.

The e-Learning programme also endeavours to provide a 'map of the territory', to provide guidance and support, and to focus development work where it will be most effective. It considers that technology should be exploited to free time from those activities that can be efficiently automated, and should be used to support innovative and creative approaches to teaching and learning.

The Advanced e-Assessment Techniques Project and its findings, along with a number of other JISC- funded projects within the e-learning programme, will contribute to the overall e-strategy. By focusing on current, advanced e-assessment practices, the project will inform JISC regarding e-assessment techniques in use, which of these can be promoted in terms of the educational value and technical robustness. The project has also been designed to provide evidence for JISC regarding the success of e-assessment projects as well as to contribute to the understanding of barriers to e-assessment.

The project, undertaken by Martin Ripley Ltd (Martin Ripley, Jeremy Tafler, Jim Ridgway, Robert Harding, Hakan Redif), utilised the skills and knowledge of experts in the e-assessment and e-learning fields. This enabled a quick and productive start to the project, covering a wide range of case studies as well as developing a lengthy catalogue of e-assessment projects in action.

## Aims and Objectives

The aim of the project was to provide JISC and its community with a review of state-of-the-art techniques in e-assessment which should be considered for application in UK HE and FE settings. This review will contribute to subsequent work to define and develop a roadmap of future e-assessment research and development activities within the JISC e-learning programme.

This JISC study was designed to build a significant body of information about who is using different techniques, the associated issues and the benefits of advanced e-assessments. The objectives of the study were:

- to develop a picture of the breadth of advanced e-assessment techniques in use around the world, and undertake a technical and educational evaluation of a selection of the most promising;

- to present short case studies describing the technical, educational and operational details of a selection of the most significant examples;
- to look at wider generic technologies and investigate how and when these might be applied to e-assessment;
- to draw conclusions about how the most relevant of these could be applied to UK FE and HE settings, and provide advice to JISC and its community about how such applications could be developed.

The project has delivered the aims and objectives by:

- identifying a range of advanced e-assessment techniques in use in HE and the wider environment that currently totals over 90 in number;
- selecting and studying in depth seven projects covering the major areas of test and assessment systems: task/test design; administration; marking, scoring and making judgements.
- obtaining information about:
  - the range of users of the selected techniques;
  - the educational and technological benefits;
  - the key issues.

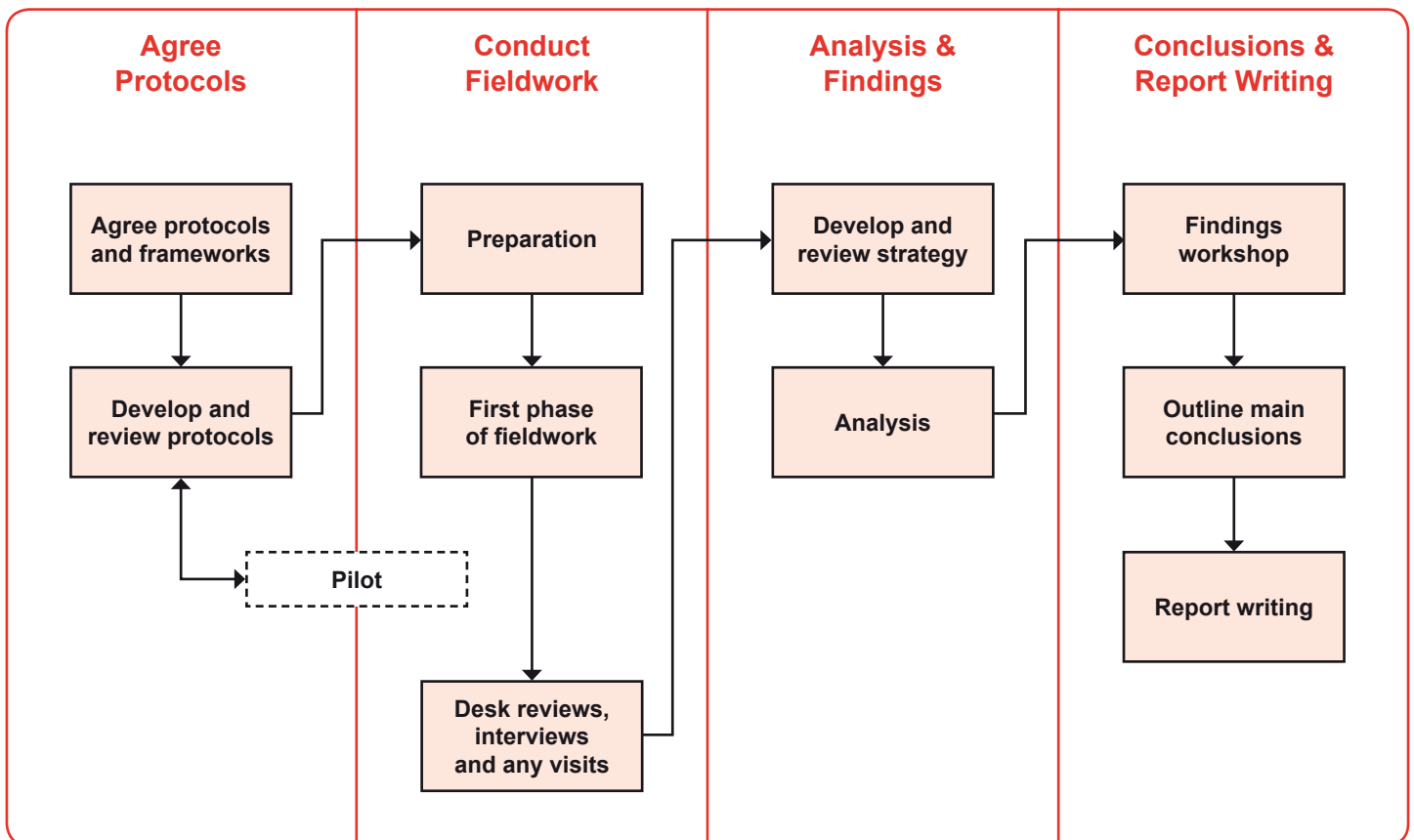
The project aims and objectives have remained consistent throughout the duration of the project, and there were no major changes made either to those objectives or to the agreed approach that was taken to achieve the objectives. The following specific questions were identified by JISC for this project.

- To what extent has the project and the related case studies contributed to increased knowledge of e-assessment?
- What were the unanticipated outcomes?
- To what extent have the project activities remained relevant to the strategic needs of the sector?
- What does JISC need to do next as a result of conclusions emerging from the project?

Responses are provided below in the Conclusions section.

## Methodology and Implementation

Our methodological approach to this project was to adopt a case study design, applying this separately for the development of the catalogue, the case studies and the technical review. Our approach, set out in the diagram below, is based on Yin's case study methods<sup>1</sup>.



<sup>1</sup> Yin, R.K. (2003) Applications of Case Study Research, Second Edition, Applied Social Research Methods Series, Vol. 34, Sage Publications. ISBN 0 7619 2551 1 <http://www.paperfree.co.uk/> Paperfree Systems Ltd – e-Portfolio system.

Through this study we documented the administrative implications and benefits of advanced e-assessment techniques. This included:

- Developing new approaches to crediting student performance;
- Developing and improving psychometric theory;
- Improving the feedback loop between assessment and learning;
- Improving accuracy;
- Speeding-up the marking stages;
- Availability of assessments on demand.

At the outset we compiled a list of advanced e-assessment techniques and identified users of those techniques. We trawled through the initial source references provided by JISC, supplementing these with our own. We reviewed existing JISC publications and consulted with JISC to ensure that we did not replicate existing information. We also used our own networks to identify additional applications. These included:

- The e-Assessment Association, which now has over 250 members including from Higher Education;
- The European Union Joint Research Centre's e-Assessment Expert network, which includes members from countries across Europe and world-wide;
- The 21st Century Learning Alliance and its associated networks, including Cisco, Oracle, Microsoft, BT and Intel;
- Our own professional network, which includes e-assessment experts in the USA, the Middle East, China and Australia;
- Virtual conference archives, including CAA and the IAEA.

The information gathered was collated into a spreadsheet. At an early stage the project realised that the initial target of 40 assessment techniques would be greatly surpassed. This therefore altered the approach of producing the catalogue from a soft/hard copy format to a web-based tool. With over 90 projects identified by the end of November 2008, the project decided to stop populating the catalogue and to start work on selecting applications/techniques to become the focus for the case studies.

It was agreed with JISC that case studies should cover advanced e-assessment issues broadly rather than in great technical depth. The advantage of this approach was considered to be two-fold: that a larger number of case studies could be produced, and the case study reports could be written for a wider range of audiences, without any particular emphasis on technical know-how.

The project considered the 90+ initial entries into the catalogue, organised them according to the e-assessment techniques that each exemplified, and provided each with a grading based on success, innovation, popularity, applicability and general uniqueness. The outcome was that JISC was presented with a shortlist of 20 possible case study candidates. A meeting was

held with JISC to discuss these candidates and to select projects which offered the potential of becoming a worthwhile and useful case study. It was agreed that some of the case studies would be themed and that, as such, they would include analysis of two or more applications of the chosen e-assessment themes. Other case studies would each provide analysis of a single-project. The final selection can be viewed in the 'Output and Results' section below.

After further research to confirm the feasibility and suitability of each nominated case study, a total of five case studies were finally agreed. Three of the five were themed. In line with the agreed research methodology, a case study research pro-forma was created and reviewed independently, before being trialled on one of the chosen case studies. As a result of the trial, a few minor amendments were made to the pro-forma and interview tool. As appropriate, site visits or telephone-based interviews were arranged. A workshop was held with JISC to review earlier drafts of analysis and to agree an overall approach for the case study reports. Subsequent working drafts of the case study write-ups were shared with JISC for their input and with an independent expert (Robert Harding) to quality assure the reports. Drafts and final versions of the case study reports were checked with and ultimately sign-off by the interviewees.

The lessons learned are set out below:

- As part of the project title, the word '*advanced*' was interpreted in a number of ways, whether *advanced* in its development/trialling and uptake, *advanced* as in the cutting edge/innovative nature of the technology, or *advanced* in terms of accepted practice, 'established' or 'in the market'. It may have helped the project in the early days of creating 'the 40+ list' to have had a better articulated definition.
- As with the word '*advanced*', similar issues were experienced with the word '*techniques*'. In this project '*techniques*' were taken to cover a wide range of features, including 'applications', 'assessment methods', 'software', 'projects' and 'concepts'.
- The attribution of 'techniques' to each of the projects in the catalogue has proved problematic. One issue is that some projects make claims (for example to be using "adaptive testing techniques") although following subsequent research this was sometimes found not to be the case. A second issue is that there is a complex matrix-relationship between projects and techniques – for example, any one project might incorporate several techniques (eg, short-text scoring, adaptive, diagnostic feedback, when ready testing), although only one or two of those techniques could be described as a dominant feature of the project.
- At the outset of the project, it was unclear whether we were listing all projects/techniques regardless of the extent to



which they contributed new understanding to e-assessment or whether the list should have been tightly focused on those that were of most interest in terms of innovation, cutting edge or proven success. In the event, it was agreed that the catalogue should contain all known projects, regardless of the extent to which they replicated previous practice or added new approaches.

- The level and amount of new, innovative developments were to an extent relatively unknown at the start of the project. From the existing research to date, it came as somewhat of a surprise and disappointment to the project team that there are not more new, innovative developments or ‘advanced’ projects being used.

## Outputs and Results

This project produced three groups of outputs:

1. An online catalogue of advanced e-assessment projects and initiatives.
2. A series of five case study reports.
3. A final report.

### Online catalogue of advanced e-assessment projects and initiatives

The project was designed to capture brief details of 40+ advanced e-assessment projects and initiatives. In the event, details of over 90 were captured. These details have now been made available through a Durham University website

<http://www.dur.ac.uk/smart.centre1/aeac>

For each project or initiative, the website lists the following information:

- Name of the project or initiative
- Institution
- Contact details of the individual or team leading the initiative
- URL
- A brief description
- Details of the advanced e-assessment techniques used in the project or initiative

In order to meet this project’s aim of identifying a range of projects using advanced e-assessment techniques, it was first necessary to create a list of techniques. The list was created initially through a brainstorm, and then supplemented throughout the project by adding techniques and descriptors as we progressed. In the event, the list of techniques contains a multiplicity of approaches: some (eg fuzzy logic/pattern matching) are “advanced” in terms of the definition of “advanced” used in this paper – that the technique is not yet commonplace in HE assessment; others are not “techniques”, but are perhaps

better seen as tools, such as PDAs; yet others describe “purposes” rather than being techniques (eg assessment impacting learning programmes).

It was not an aim of this project to create a glossary of advanced e-assessment techniques. In the event, it was found that simply listing advanced e-assessment techniques is more complex than had been anticipated. However, the following list could be developed further into a more rigorously defined set of techniques. It should be pointed out, of course, that any single project or initiative typically incorporated more than one technique.

### Access

The use of technology designed to enable a greater range of students to understand and be able to take part in an assessment

### Accuracy of complex inputs

The use of technology to measure the technical accuracy of students’ responses, such as in foreign language translation exercises

### Adaptive testing

Assessments which deploy algorithms to analyse student responses during the test session and to vary the selection of items presented to the student accordingly

### Assessment impacting learning programmes

The use of technology to create dynamic links between the assessment and learning programmes

### Assessment on demand

The use of technology to make assessments and tests available to students at a date and time of the student’s choice. This technique was also used to refer to assessment systems which gave the teacher or tutor choice over the timing of assessment (i.e. the timing decision was made as locally as possible.)

### Associated services (e.g. plagiarism detection)

The use of technology to provide services allied to assessment and testing; sometimes not available in pre-technology forms

### Complex mathematical responses

Analysis of student responses which include mathematical expressions which appear in many forms and which might include algebraic or more complex formula-based expressions

### Data mining

The use of technology to support data mining techniques.

### Diagrams

The use of technology to enable the presentation of diagrams, graphs and related approaches in assessments, as well as enabling students to manipulate, draw and/or label diagrams by way of response

### **Expert judgement/support systems**

An application that uses a knowledge base of human expertise and algorithms to emulate decisions that would be made by a human

### **Extended text**

The use of technology to support input and/or scoring of prose responses in excess of 20 words or so.

### **Feedback**

The provision of information and analyses for tutors, teachers and students based on individuals' and groups' performance on tests and assessments

### **Formative**

The provision of information and analyses for students where the primary purpose of this is to inform decisions about next steps in learning

### **Fuzzy logic/pattern matching**

The use of string matching computer techniques to compare students' test responses to predicted and expected answers

### **Group work**

The use of technology to set students group assessments, and to use technology to collect information about the performance of the group and, sometimes, to score individual's contributions to the overall group performance

### **Higher order skills**

The use of technology to assess students higher order skills, including process skills (e.g. problem solving) and behavioural attributes (e.g. persistence and logic)

### **Interactive skills**

The use of technology to assess student's ICT skills

### **Knowledge**

Assessment of domain knowledge, potentially using a variety of item types

### **Marking combining human/computer**

The use of technology to filter responses which can be automatically marked, presenting to human markers all responses that were not predicted or that cannot be automatically scored for other reasons

### **Mobile phones**

The use of mobile phones to present assessment tasks and/or to collect student responses using the phone as an input device

### **MUVes & Virtual Worlds**

The use of Multiple User Virtual Environments or other types of Virtual World to conduct assessments. The virtual worlds

could be used as a vehicle for presenting a range of types of assessment, including those which are designed to assess performance in using virtual world functionality and interactivity

### **New approaches to student performance**

The use of technology to design and present measurements of student performance not undertaken in paper-based or more traditional assessment. These innovations could be relatively technically and psychometrically straightforward (eg using video to present assessment stimulus material) or significantly more sophisticated (use of technology to assess group work, for example)

### **PDA's**

The use of PDA's to present assessment tasks and/or to collect student responses using the PDA as an input device. In some instances, the PDA might be used as the tutor's assessment handbook, reminding the tutor of the assessment rubrics and requirements

### **Peer assessment**

The use of technology to collect peer-to-peer assessment judgements, possibly including the use of technology to present the assessment evidence

### **Practical skills**

The use of technology to assess motor skills, physical skills, performance evidence and other types of practical skills

### **Process**

The use of technology to collect evidence of student's process performances, such as methods used to solve problems

### **Psychometric theory**

The design of new and adaptive forms of performance measurement, leading to advances in psychometric theory not based on linear, item-based and paper-based forms of testing

### **Short free text**

The use of technology to capture and score students' short free text responses to test items.

### **Speed**

The assessment of time-based aspects of students' performance, including the speed of their responses

### **Training**

The use of technology to derive training-needs analysis, for instance by evaluating gaps and weaknesses in students' performance.

### **Voice/performance skills**

The use of technology to capture voice (i.e. spoken word) evidence of performance

## Web2 – flickr, wikis

The use of social technologies as a platform for presenting assessment tasks and/or capturing evidence of student performance

## VLEs

The use of Virtual Learning Environments to launch assessments of any type

## Case study reports

On reviewing the catalogue of e-assessment projects to select some to study in greater depth through case studies, it was agreed with JISC that two types of case study would be research. The first type would look broadly at a particular

e-assessment technique, reviewing a number of initiatives to provide analyses of issues underpinning that technique. These thematic case studies examined Short Text, Foreign Language Textual and Spoken and Storage/databanks. The remaining case study reports take as their focus the specific e-assessment project.

The table below summarises the case studies, also providing information about the name of the institution, lead contact and research completed.

These five case study reports have been published as separate and stand-alone documents. They are available at <http://www.dur.ac.uk/smart.centre1/aeac> In brief, the case study reports cover the following content.

Case Study Report and its focus	Institution(s)/company	Contact	Research completed
<b>AsTTLE</b> Adaptive testing	University of Auckland	John Hattie	Telephone, email and desk-based research
<b>Assessment 21</b> Combining human/computer marking	University of Manchester	Dr Mary McGee	Email and desk-based research and site visit on 2 March 2009
<b>Primum</b> Higher order skills	National Board of Medical Examiners (NBME) Philadelphia USA	Brian Clauser	Telephone, email and desk-based research
<b>Foreign language textual and spoken</b>	LISC – University of Kent	Alison Fowler	Email and desk-based research and site visit on 2 March 2009
	Pearson Versant	Jared Bernstein and Alistair Van Moere	Telephone, email and desk-based research
<b>Short text</b>	Automark, University of Dundee	Walter Williamson, John McEwen	Email and desk-based research and site visit on 20 January 2009

## Adaptive testing

The focus of this case study is AsTTle. AsTTle is an e-assessment application, the main function of which is to provide users (tutors or teachers) with a system which enables them to specify and select test items to configure a test. The selection might be predicated on a focus on a particular curriculum domain, for example, or on the demand of the test items. The benefits of such a system derive from the ability to provide teachers with access to large, high-quality and calibrated item banks whilst also providing them with the flexibility to create test forms which closely reflect their teaching. Interestingly, AsTTle is an application developed in HE (in New Zealand), but providing services to school teachers.

## Higher Order Skills

Some advocates of e-assessment point to the potential of computers to support simulation and scenario-based assessment. There are few examples of this category of e-assessment being developed successfully, especially not in high stakes testing contexts. Primum is an exception. It provides an assessment of trainee medical practitioners' ability in making medical diagnoses when presented with a fictitious patient exhibiting a number of symptoms. This automated assessment has been designed to provide an authentic and reliable assessment at a price that compares favourably with the alternative – human scored evaluation at patients' bedsides.

## Combining human and computer marking

Few e-assessment systems are designed to drive efficient marking by automatically sorting and classifying students' responses, then directing the sorted listing of responses to human markers. Assessment21 does. The benefit of such an approach is that it delivers more efficient marking (a human marker can browse sorted responses, applying one judgement to multiple, identical responses) as well as supporting greater consistency in marking (a marker views all similarly constructed responses at one time and applies a single, common judgement to all of those similar responses).

## Automatic scoring of foreign language textual and spoken response

Computers have only limited ability to analyse and score the grammatical complexity of written language, and accuracy and fluency in the spoken word. Two e-assessment applications which do are LISC and Versant. LISC provides a detailed analysis of the grammatical accuracy of sentences translated into a second language by students, supporting students in their understanding and application of grammatical rules. Versant provides an automated spoken language proficiency assessment

with applications as diverse as Dutch government citizenship tests, university entrance fluency tests, and air-traffic controllers' proficiency tests. By designing assessments of highly focused skill domains, both of these e-assessments deliver effective assessments of skills that are costly to provide in a human-scored system.

## Automatic scoring of constructed, short text responses

There are a number of e-assessment products which score short-text responses. This case study provides a survey of current practice, with a particular focus on two systems - Automark and IAT. The strengths and limitations of short-test scoring are well understood and, although not a panacea, short-text scoring is a particularly strong aspect of e-assessment. Short-text scoring systems provide the benefits of using constructed responses items (which can provide valid assessments of students' understanding and application of knowledge) and provide the technology-driven benefits of efficiency and accuracy.

## Findings

Through the process of compiling the Advanced e-Assessment Techniques catalogue, we arrived at a number of findings.

### F1

#### Gaps in the range of techniques in use

The project team found that there was very little evidence of some of the more advanced, Web2 technologies being used for assessment purposes. Although social technologies, virtual worlds and learning environments are to be found throughout higher education, we found little evidence of assessments being conducted in these environments.

### F2

#### Few recent additions

The project team found that the majority of projects listed within the catalogue were reasonably well-known, with far fewer than expected new developments identified. It might be that the methods used to seek out projects were not sufficiently refined to find these newer projects: however, the fact remains that the e-assessment community(ies) does not know of many new up-and-coming e-assessment initiatives.

### F3

#### The isolated nature of some projects

The project team found a number of projects which, although encompassing significant e-assessment expertise and relevant technical developments, remain essentially isolated and in some cases are continuing to be led by a single person. This lack of 'spread' of technologies and approaches is reflective of the earliest stages of an innovation cycle, in which the developments are led by a very small minority of early-adopters who are often the experts and gurus.

The purpose of the case study reports is to provide detailed information to the JISC community about specific e-assessment developments. As such, the studies have been designed as stand-alone digests, rather being designed to inform strategic considerations. However, the following findings emerge from our case study research.

**F4**

#### **The isolated nature of some developments**

In parallel with finding F3 above, the case studies provide evidence of the difficulties in 'growing' the e-assessment application for wider use. The LISC study is an example of a highly reliable and technically specific e-assessment development which has been driven by a single person, which does not have sufficient funding or resource to take it to a next level of operation, and which has almost become caught in a set of circumstances that threaten to keep it small scale. Primum, Versant and short-text are all examples of e-assessment developments which have achieved a very high level of technical robustness, with associated depth of research and with considerable surrounding expertise. And yet these technologies have not won wider adoption.

**F5**

#### **The human factor and the need to demonstrate business efficiencies**

Ultimately, e-assessment technologies are only likely to be adopted if tutors and students see the benefits. Advocates of e-assessment seek to point to three areas of potential benefits of e-assessment:

- Enhancements to the validity of assessment
- Business improvements – e.g., better diagnostic information, more flexibility over the timing of assessment, quicker marking and results notification
- Supplier benefits – e.g., cheaper to run and a less risk-prone business

It is clear from several of the case studies that delivering at least some of the above benefits is a necessary but not sufficient condition for wider adoption. Despite significant claims of being able to provide these types of benefit, a critical ingredient for the successful wider adoption of e-assessment projects is an ability to overcome concerns and doubts about technology (i.e. a human perception factor) and an ability to demonstrate benefits to the tutor (usually in the form of time savings).

**F6**

#### **Links across sectors**

Several of the case studies related to e-assessment developments which originated in HE, but which have grown through adoption in other educational and training sectors. LISC was developed at the University of Kent; all of its current users are secondary school-based. AsTTle is run from the University of Auckland, providing e-assessment tools for schools. Versant technology grew out of highly specialised research and design at the University of Cambridge, and is now providing commercial success to Pearson. In selecting case studies with

JISC, the project team also identified the fact that some e-assessment techniques are not being pursued within HE. The use of item banks, to share the highest quality test items across institutions is a practice found extensively in the UK schools sector, but not at all to the same extent within HE. Both at a practical and at a strategic level, therefore, the case studies point to the benefits of cross-sector collaboration in e-assessment.

Conclusions based on the above findings are presented later in this report.

## **Outcomes and Impact**

This JISC study was designed to build a significant body of information about who is using different techniques, the associated issues and the benefits of advanced e-assessments.

The objectives of the study were:

- To develop a picture of the breadth of advanced e-assessment techniques in use around the world, and undertake a technical and educational evaluation of a selection of the most promising;
- To present short case studies describing the technical, educational and operational details of a selection of the most significant examples;
- To look at wider generic technologies and investigate how and when these might be applied to e-assessment;
- To draw conclusions about how the most relevant of these could be applied to UK FE and HE settings, and provide advice to JISC and its community about how such application could be developed.

The project has delivered the aims and objectives by:

- identifying a range of advanced e-assessment techniques in use in HE and the wider environment that currently totals over 90 in number;
- selecting and studying in depth five case studies covering the major areas of test and assessment systems: task/ test design; administration; marking, scoring and making judgements.
- obtaining information about:
  - the range of users of the selected techniques;
  - the educational and technological benefits;
  - the key issues.

The evidence gathered throughout this project has also led to findings with wider implications for JISC. HE institutions differ a great deal; subject disciplines view knowledge in different ways. There exists a variety of views on the purposes of assessment in different subjects from professional accreditation, through comparing standards in different institutions, to putting students

in a particular cohort into rank order. Assessment serves a great many functions: it can define what is worth knowing, and can steer teaching and student learning. From this perspective, e-assessment projects will often struggle to find common ground and users across a range of institutions.

In a fast developing field, one would hope to see developments at a number of levels – in particular, a ‘bleeding edge’ where enthusiasts pioneer novel approaches, and a fat tail where techniques of proven effectiveness are accepted and absorbed into the daily practices of institutions. A striking feature of this review is the lack of take-up of e-assessment in HE, despite the extensive development work that has taken place, and strong advocacy.

A review of the literature on innovation is inappropriate here, but some summary remarks can illuminate the reasons for the current situation, and can suggest ways forward.

Ridgway (2003) argued that users are likely to adopt some new way of working if:

- It solves a problem they know they have got
- It makes life more interesting and more fun
- It makes life easier
- It has a measure of social approval.

It is easy to identify circumstances where e-assessment would fail on every category. Consider the application of automated essay marking to a specialist module taught by two tutors on a philosophy course. Currently, the tutors mark all the essays jointly. The work load is not high, and the task is interesting – so there is no problem to be solved. Automated assessment would remove an interesting part of the teaching, and would reduce job satisfaction because the tutors no longer read essays on a topic central to their research interests, and they lose the opportunity for discussions triggered by student conceptions and misconceptions. The task of assembling enough scripts to facilitate the automated scoring scheme is probably impossible. Further, an important source of evidence for revising teaching on the course has been lost. Students would complain vehemently about the lack of relevant feedback, and colleagues would be outraged by the conceptions of knowledge embedded in the procedure.

However, it is possible to identify some fundamental questions and issues which new designers of e-assessment solutions might wish to consider. (A wider list of ‘questions to use when making decisions about investing resources in e-assessment’ is provided as Appendix Three.)

### **What (if any) problems are you trying to solve with the new system?**

- Efficient, frequent assessment of large numbers of students (and associated consequences for tutors in terms of marking workloads)?
- Identification of and support for students’ knowledge and understanding? (e.g., the design of LISC to support student’s grammatical understanding in other languages).

### **How will the proposed system improve teaching?**

- By providing more authentic assessment? (e.g. of difficult-to-assess goals such as process skills and student reflection (e.g. Pimum’s assessment of trainee medical practitioners’ diagnostic skills, and Pearson Versant’s assessment of the spoken word.)
- By improving teachers’ pedagogic knowledge? (e.g. by exposing teachers to student misconceptions by assembling all the answers to the same short question (e.g. Dundee))?
- By providing a map of student performance on different parts of a course, as reflected by different questions
- By removing drudgery from assessment? (e.g. Assessment21)

### **How will the proposed system improve learning?**

- By providing feedback that is both timely (so that early mistakes do not develop into malknowledge (e.g. LISC), and useful (so that students are given guidance on how to improve their performances)?
- By facilitating student autonomy and self regulation? (e.g. LISC)

### **How will the proposed system improve the assessment system?**

- By drawing attention to the importance of good design in the creation of assessment tasks?
- By increasing inter-marker reliability?
- By raising awareness of key issues in assessment with academic staff (alignment, validity, reliability)?
- Where appropriate (especially with large class sizes), widespread adoption of techniques that can help improve the design of assessment, such as routine use of psychometric techniques - examining various test reliabilities, means and variances of different components that contribute to students’ overall grades, and using techniques associated with Item Response Theory? (e.g. Assessment21)
- By increasing the variety of assessment, or time spent by staff giving feedback on some aspects of students’ work, if time is saved on routine aspects of marking? (e.g. LISC)
- By increasing fairness?

## How will the proposed system improve administrative systems?

- By providing timely information? (e.g. LISC)

## How will the proposed system address core goals of the university?

- By supporting professional dialogues between staff?

**F7**

### The need to be clear about benefits

Linked to Finding F5, it became clear through the case study analysis that none of the case studies addressed very many of these questions and issues. Some of the questions and issues were not covered by any of the case studies. A more robust framework for analysing benefits to a wider range of users within HE institutions, and earlier efforts to secure stakeholders' involvement in and ownership of e-assessment developments might be helpful.

It is possible also to go a significant step further. Generally, we tend to look at e-assessment projects to find interesting, useful and clever new techniques that they offer. A different perspective is to analyse the assessment and related problems being experienced by HE institutions and to identify ways in which technology might be able to address these to a greater or lesser degree.

For example, there are a number of serious concerns about current assessment practices in HE, which include:

- Low inter-marker reliability
- Comparability of qualifications across universities
- The richness and variety of assessment practices (essentially that a monoculture of essays is not acceptable as a way to assess attainment)
- Low quality feedback to students (in terms of timeliness, or specificity)
- High levels of self-reported plagiarism

Related to this, there are a number of current concerns expressed about e-assessment – whether based on information and evidence, or as perceptions:

- Curriculum narrowing by assessing things that are easy to assess (e.g. MC assessment of factual knowledge), or by an imposed uniformity across HE institutions
- Rewarding irrelevant skills (e.g. MC can reward the skill of choosing the right answer from the alternatives, as well as knowing it)
- Claims that seem conceptually impossible (computer-based assessment of essays)
- Poor feedback to students, in the form of grades without information on how to improve

- Loss of teacher autonomy; dependence on external support
- Perceived loss of ownership by teachers, and student perceptions that a course is a mass product, not one designed with them in mind
- Technical problems (resource implications, reliability, access issues, transfer across computer platforms, vulnerability to viral attacks, and hacking)

## Conclusions & Recommendations

This study has found that there are some impressive examples of e-assessment, but rather little evidence of widespread use (the OU, and Manchester ABC provide notable exceptions). The field seems immature: there is a large gap behind the 'bleeding edge' and the day-to-day life of departments, and there are scant instances of e-assessment innovations being adopted beyond the initial, small circles of early adopters. There is little evidence of large scale adoption by prestigious universities.

In order to address this conclusion, and based on the findings outlined above, this report makes the following recommendations.

**R1**

Best practice in assessment should figure more fully in HEFCE's TQA activities. Obvious aspects to review include: validity and reliability; the frequency and quality of the formative feedback students receive; self-assessment that can foster student autonomy; and the speed and accuracy of feedback. This increased emphasis is likely to encourage more use of e-assessment, and the development of more ambitious e-assessment systems aligned closely with academic values.

**R2**

There is a need to survey current assessment practices (especially on large courses), to understand the seriousness of concerns about alignment, validity, reliability and plagiarism.

**R3**

Work could be commissioned to extend and maintain the database of e-assessments created here. The Field Tested Learning and Assessment Guide (FLAG) project created by the National Institute for Science Education in the USA provides an example).

**R4**

There is a need to promote styles of work focussed around assessment that empower students to take more control of their learning. This is most appropriate in first year courses. Discussions should centre around what the student knows, does not know, and on strengths and gaps. Learning to self-assess, and developing metacognitive skills of learning to learn should be important foci for learning. Self-test facilities for students (e.g. Dundee Medical School) should be implemented and studied in action.

**R5**

JISC might consider limited funding to stimulate prototype development of e-assessment using some of the techniques for which this project reported little or no assessment usage – such as Web2, MULEs, peer-to-peer assessment. In addition, Appendix Two sets out a number of possible e-assessment situation analyses designed to offer focal points for leading e-assessment solutions to the next stage of wider adoption.

**R6**

JISC might consider pro-active “marketing” of some of the projects studied in this report – LISC, Assessment21 and short text scoring are projects that provide benefits which would be of relevance to many HE institutions.

There are other implications for licensing practitioners from non-EU countries. E-assessment should be used to ensure that professionals reach adequate levels of proficiency in language and mathematics (where relevant). A more satisfactory situation (from the viewpoint of safeguarding UK citizens) would be an extensive set of assessments with a strong e-assessment component, exemplified by PRIMUM.

## Implications for the future

The review has found many innovative and well researched examples of e-assessment techniques. However, there is little evidence that e-assessment is having a substantial and positive impact on learning and teaching in UK universities. Current cultural and administrative practices in HE often mitigate against the propagation of successful initiatives.

This has a number of important implications for the future of UK HE, in terms both of the future of UK HE, and in terms of our role and status in Europe. The implications for the future of UK HE led to the commissioning of this work, and so will not be elaborated here. Implications for UK HE in Europe bear further analysis.

The Bologna accords set out to align academic degree standards across Europe. This will have an impact on the whole degree accreditation process. Aspects of e-assessment will be critical to this process. There is scope for the UK to take a central leadership role here. There are obvious intellectual and economic benefits in being seen as the major providers of academic benchmarks for pan-European HE. Failure to do so could lead to some marginalization of UK qualifications. An important implication of this review is that resources should be devoted to e-assessment designed to promote pan-European alignment of qualifications. An important corollary to this is the likely impact of European attitudes to education on e-assessment, notably to emphasise a broader range of skills, such as performance on extended tasks, and the ability to work with others.



## References

Assessment21. Available at [www.assessment21.com/resources.html](http://www.assessment21.com/resources.html)

asTTle. Available at <http://www.tki.org.nz/r/asttle/> and [http://www.tki.org.nz/r/asttle/e-asttle/faq\\_e.php#interpreting-reports6](http://www.tki.org.nz/r/asttle/e-asttle/faq_e.php#interpreting-reports6)

Balogh, Jennifer; Bernstein, Jared; Suzuki, Masa; Subbarayan, Pradeep; Lennig, Matthew. *Automatically Scored Spoken Language Tests for Air Traffic Controllers and Pilots*. Harcourt Assessment, Menlo Park, California, USA. Available from <http://www.icao.int/trainair/meetings/gtc10/pivp5.pdf>

Braun, Henry et al, Rule-based methods for automatic scoring: application in a licensing context. In Williamson, Misley and Bejar. *Automated Scoring of complex tasks in Computer-Based Testing*, Mahwah New Jersey 2006

Butcher, P. (2008). Online assessment at the Open University using open source software: Moodle, OpenMark and more. 12th International CAA Conference, Loughborough, UK. [www.caaconference.com/pastConferences/2008/proceedings/index.asp](http://www.caaconference.com/pastConferences/2008/proceedings/index.asp)

Fowler, Alison. *LISC: web-based CALL for Language Teaching*, University of Kent

Hattie, John. *Large-scale Assessment of Student Competencies*. University of Auckland, New Zealand. Paper presented as part of the Symposium: Working in Today's World of Testing and Measurement: Required Knowledge and Skills (Joint ITC/CPTA Symposium); 26th International Congress of Applied Psychology; July 16-21, 2006, Athens, Greece

Hincks, Rebecca, *Using speech recognition to evaluate skills in spoken English*. Lund University, Dept. of Linguistics, Working Papers 49 (2001), 58–61. Available from <http://conference.sol.lu.se/fonetik2001/proceedings/bidrag15.pdf>

HTK and the Hidden Markov Model, see <http://htk.eng.cam.ac.uk/>

Intel, Microsoft, and Cisco Education Taskforce. *Transforming Education: Assessing and Teaching the Skills Needed in the 21st Century. A Call to Action*. Available at [ ]

Jordan, Sally and Mitchell, Tom (2009). *E-assessment for learning? The potential of short free-text questions with tailored feedback*. *British Journal of Educational Technology*, 40, 2, pp. 371-385

Leacock, C and Chodorov, M. (2003). C-rater: Automated Scoring of short answer Questions. *Computers and Humanities*, 37, 4, 389-405. See also [http://www.ets.org/Media/Research/pdf/erater\\_examens\\_leacock.pdf](http://www.ets.org/Media/Research/pdf/erater_examens_leacock.pdf)

LISC Homepage <http://www.cs.kent.ac.uk/people/staff/amlf/CALL/HTML/menu.html>

Margolis, Melissa J and Clauser, Brian E. A regression-based procedure for automated scoring of a complex medical performance assessment. In Williamson, Misley and Bejar. *Automated Scoring of complex tasks in Computer-Based Testing*, Mahwah New Jersey 2006.

*Primum Computer-Based Case Simulations (CCS) - Tutorial and Practice Cases*.

Available at <http://www.usmle.org/Orientation/2009/menu.html>. See also [www.usmle.org](http://www.usmle.org)  
[www.usmle.org/Examinations/step3/step3.html](http://www.usmle.org/Examinations/step3/step3.html)  
[www.usmle.org/Examinations/practice\\_materials.html](http://www.usmle.org/Examinations/practice_materials.html)

Pulman, S. & Sukkarieh, J. (2005) Automatic Short Answer Marking. Proceedings of the 2nd Workshop on Building Educational Applications Using NLP, Ann Arbor, June 2005. <http://www.comlab.oxford.ac.uk/people/publications/date/Stephen.Pulman.html>

Ridgway, Jim (2003) The evaluation of complex competencies. Invited keynote at the V Seminario de la Redu (focussed on the responses of Spanish Universities to the Bologna accord), Madrid. (paper available from author).

Ripley, Martin *Transformational computer-based testing*. In Scheuermann, Friedrich and Björnsson Julius (Eds.). *New Approaches to Skills Assessment and Implications for Large-scale Testing: The Transition to Computer-Based Assessment*. JRC, Ispra Italy, 2009

Ross, S.M., Jordan, S.E & Butcher, P.G. (2006). Online instantaneous and targeted feedback for remote learners. In C. Bryan & K.V. Clegg, K.V. (Eds), *Innovative assessment in higher education* (pp. 123–131). London: Routledge.

Scheuermann, Friedrich and Björnsson Julius (Eds.). *New Approaches to Skills Assessment and Implications for Large-scale Testing: The Transition to Computer-Based Assessment*. JRC, Ispra Italy, 2009

Sukkarieh, J. Z. , Pulman, S. G. and Raikes, N. (2003) Auto-marking: using computational linguistics to score short free text responses. Available at <http://www.clg.ox.ac.uk/pulman/pdppapers/AUTOMARKING2.htm>

Young, Steve; Kershaw, Dan; Odell, Julian; Ollason, Dave; Valtchev, Valtcho; Woodland, Phil. *The HTK Handbook*. Available from <http://nesl.ee.ucla.edu/projects/ibadge/docs/ASR/htk/htkbook.pdf>

Williamson, Misley and Bejar. *Automated Scoring of complex tasks in Computer-Based Testing*, Mahwah New Jersey 2006.

## Appendix One: Glossary

### **Advanced**

*“Advanced”* – as used in “Advanced eAssessment Techniques” – refers to techniques that are used in isolated or restricted domains, and which have successfully applied technology to create an assessment tool. “Advanced” does not necessarily imply “newness”.

### **Techniques**

*“Techniques”* covers a wide range of features, including ‘applications’, ‘assessment methods’, ‘software’, ‘projects’ and ‘concepts’

## Appendix Two: Extended e-assessment situation proposals

The following four situation sketches are designed to illustrate aspects of assessment practice in HE institutions where e-assessment projects could be commissioned to deliver specified benefits. This approach is attractive because the assessment issues are widely understood to be issues. The approach therefore would be to use e-assessment to solve known problems, as opposed to “marketing” e-assessment on the technology benefits and enhancements.

### Situation 1: E-assessment targeted on known problems

Possible areas for a JISC e-assessment focus:

#### Student cheating areas

This is a major problem in HE (see e.g. Smith and Ridgway, 2008). Technical approaches can never provide a complete solution, but they can help. Examples include:

- the plagiarism checker Turnitin, that provides information on the likely source of each student’s essay;
- Randomisation of data values within standard assessment tasks, to ensure everyone gets the same task, but with a unique data set;
- The increasing use of essay mills such as bestessays where students can purchase custom-written essays shows a need for universities to keep a database of every essay written by each student, and submit them all to stylistic analyses of the sort used in literature studies. Anomalous essays can be investigated further.

#### Inconsistent marking by staff

This is a matter of concern on large courses where a number of staff share marking, and on courses that lead to national accreditation, where consistency of awarding practices across institutions is important.

### Situation 2: E-assessment targeted improving interactivity and motivation

Possible areas for a JISC e-assessment focus:

#### Creating innovative activities

Some academics find this inherently exciting and enjoyable.

#### Presenting students with novel challenges

This can enliven teaching and learning, and forces students to target their learning on deep understanding, rather than on examination techniques

### Situation 3: E-assessment targeted on delivering efficiency benefits

Possible areas for a JISC e-assessment focus:

#### **Large scale assessment on topics where knowledge is uncontested**

Possible examples are courses where factual knowledge and robust procedural knowledge has to be acquired, where the high throughput of students and reuse of questions makes the time spent on generating computer mark schemes worthwhile. Current European initiatives to standardise the assessment of language competence provide one example; in the context of ICT competence, the European Driving Licence, and Microsoft accreditation provide others.

#### **'Unintelligent' support for essay marking**

Such as highlighting key words, and using 'surface feature' measures to triangulate human markers (an early study should explore the extent to which such systems will discriminate against non-native speakers).

#### **Propagating marker decisions across a set of short or one word answers**

This can save time, because markers do not have to read every script

### Situation 4: E-assessment with a measure of social approval

Possible areas for a JISC e-assessment focus:

#### **Activities that are seen to be valuable, but that are impossible without technology**

Possible examples are activities where students use the web to access data, or use ICT-based tools such as spreadsheets to build models of situations

#### **Activities that demonstrate new 'core skills'**

For example, requiring students to demonstrate fluency in Web 2.0 activities.

#### **Situations where process skills must be assessed**

There are situations where ICT is the only possible solution to authentic assessment, as student numbers rise and the alternative (observation of actual practice in (say) Accident and Emergency Departments) provides a small, erratic set of experiences for students. Opportunity is a problem, standardisation is a problem (because student actions in live settings produce new situations that require new responses; and standardisation becomes very difficult) e.g. PRIMUM

## Appendix Three:

### Questions for individuals and departments to use when making decisions about investing resources in e-assessment

#### Impact on teaching and learning:

- Does it solve a problem you know you have got?
- What are the potential benefits and downsides?
- Will it make life more interesting and more fun for everyone?
- Will it make life easier, overall?
- Do key constituencies (students, colleagues in the department, in related departments in other universities, in your own university, external examiners, university administration and management) believe that this is the right direction of travel?
- What are the potential benefits for teaching and learning?
- Is it easy to re-use tasks/tests?
- Is it easy to add new content, change what is there, etc, choose what to present?
- Can it be used for other purposes?
- Is it robust?
- Is it vulnerable to plagiarism? Are there facilities for detecting cheating and plagiarism?

#### Systems issues:

- What are the implications for the assessment system as a whole?

#### Barriers:

- Are there any obvious barriers to widespread implementation?
- How flexible is the proposed system?
- How much/what sort of staff training is required?
- Training to use the system
- Authoring: creation of items and tests and modifications
- On-line marking time

#### Set up costs:

- What does it take to get started (in terms of expertise, set up time)?
- How easy is it to set up for each wave of assessment?
- How much time is involved in maintaining the working system?

#### Impact on teaching and learning:

- Is there evidence from other institutions?
- What are the potential benefits, and potential threats?

#### Risks:

- What are the risks?
- What are the costs?
- What are the threats to the organisation?

#### Market issues:

- How big is the potential demand for this sort of assessment?
- What is its strategic importance?
- Are student numbers expected to rise in this area?
- Are there issues of quality assurance that need to be resolved

#### What is the evidence base?

- Where has this application been used successfully?
- How well does that context fit your own context?
- What is the extent of the current use of the application being considered?
- What evidence is there on desirable and undesirable impact?

#### Adoption:

- To what extent has it rolled out within the originating institution?
- Has use extended beyond the originating institution to other institutions?
- Are further generalisations plausible?

#### Dilution and corruption:

- What will systems look like when they are being used by people who did not invent them?





# Review of Advanced e-Assessment Techniques (RAeAT) Final Report