User Experiences of Choosing and Implementing Web-Based Computer Aided Assessment

Andy Crofts*, Maureen Foster✦ Mhairi McAlpine, Roger Rist, and Nils Tomes
Learning Technology Centre, Heriot Watt University, Edinburgh
* Web4Test Ltd, Edinburgh
✦Department of Mathematics, Heriot Watt University, Edinburgh

Abstract

This paper aims to give an insight into the decision making process followed in identifying how web based assessment tools can complement a broader assessment strategy. Building on a number of evaluation trials carried out with Scottish universities, a framework has been developed which helps institutions identify the key criteria which would enable them to form a strategy for their own organisation, and use these to select appropriate web based technologies.

Heriot Watt University, in partnership with the Universities of Aberdeen, Dundee and Napier, has been trialling the WebTest assessment engine (http://www.ltc.hw.ac.uk/lts/scaan/) and evaluating how well it meets the needs of a diverse group of educational practitioners within different institutional settings. Members of key identified user groups (students, lecturers and administrators) from the departments trialling the assessment engine participated in workshop sessions designed to identify their requirements from a CAA system and to re-evaluate these in the light of practical implementations of web-based assessments.

This paper draws on the experience of Heriot Watt University, as part of the Scottish Computer Assisted Assessment Network (SCAAN) consortium, which has been trialling three web-based CAA engines at a number of Scottish Universities. It documents both the process of choice and of implementation and highlights a strategy for departmental and institutional adoption of an appropriate engine.

Introduction

The Scottish Computer Assisted Assessment Network was established in 1998 to raise awareness within the Scottish Higher Education community to the recent advances in Computer Assisted Assessment. The Network is a consortium of three Universities (Strathclyde, Glasgow and Heriot-Watt) each of which uses a different web-based assessment engine with their students. A fuller description of the work of the SCAAN project can be found in Sclater and Howie (2000).

SCAAN is currently trialling the three engines (Miranda, WebTest and Triads) within a number of institutions, including Glasgow Caledonian University, Aberdeen University, St Andrews University, Napier University, Paisley University and Edinburgh University, in a wide range of departments from Genetics to Nursing. The aim of these trials is not only to allow institutions

access to assessment software at no cost, but also to evaluate how well currently available assessment engines being developed in Scottish Higher Education establishments meets their needs.

The initial web based assessment approaches development at the University fulfilled a pragmatic need. The initial impetus came in the mid-1990s from the requirement to support distance learning mathematics students, who were geographically dispersed and who had access to variable computing infrastructure. The educational need was to provide learners with sets of example mathematical problems to reinforce their understanding of areas of the curriculum, build up strategic skills in tackling different styles of mathematical problems, and to give the learners the feedback which they needed to help them direct their studies effectively (Foster and Crofts, 1999). The learners had a great appetite for sheets of mathematical problems, which was hard to satisfy in a timely way by conventional means. Although they used tutorial support through telematics-based conferencing, students liked working through additional examples in their own time: when they completed the supplied examples they would, like Oliver Twist, demand some more. Providing problem-generators or item banks seemed a pragmatic way forwards. However, the students had access to a diverse computing infrastructure with limited technical backup, as they studied from a range of local study centres or from home. The development, maintenance and distribution costs of providing a desktop application were prohibitively high, and the web provided an obvious avenue to delivering cost-effective services.

The WebTest assessment engine had a number of key features which addressed these initial user requirements (Crofts, 1999): the ability to display and judge a broad range of mathematical problem types, the ability to include variable parameters within questions so that sets of analogous examples could be generated by the user. From the organisational point of view, the system was designed to be proof against technology changes by defining assessment questions in a manipulatable mark-up language, and use of the cross-platform web browser reduced maintenance and support needs. From the users point of view, it provided the required baseline services of being reliable, delivering accurate judgements, being simple to use, and sparing of band-width.

Developing on from this, the WebTest system proved its value in supporting both on-campus and distance learning mathematics students and the system's range of question types was extended to meet the needs of students studying other disciplines, with the system being implemented in all faculties of Heriot-Watt University. To support this growth in use of web-based assessment through planned development, the WebTest service providers set out to identify user requirements for web based assessment which could be used as part of a method to identify and prioritise key criteria. This had two main aims: to provide a method for new users to identify whether the available web based assessment tools would meet their requirements, and to identify priority areas for development of system functionality.

The use of this web-based computer assisted assessment offered teachers and tutors some organisational advantages in monitoring their students' progress, and evaluating their course materials, and their requirements of the

kinds of information which they needed grew. As part of service delivery, the WebTest group recorded student and teacher comments, and monitored new user requirements, building up a expanded list of user requirements. Formal evaluations with students and teachers were also carried out, which focussed on educational, organisational and technical needs (Mochrie,1998), and these evaluations were extended beyond the host institution through the Scottish Computer Assisted Assessment Network and with institutions elsewhere in the UK and Europe.

In 1998, it seemed timely to re-consult users on their requirements of a computer aided assessment system, as the diversity of the user group had grown, teaching and learning methods were evolving, and delivery technologies continued to develop.

The evaluation of learning technology developments is a rather under-researched area. Most evaluative studies have a narrow focus, such as the reliability of the results (Russell, 1999), the learning strategies adopted by users (Beishuizen, 1999) and the ethical implications (Bennington, 1999). The focus of this evaluation however was to assess how well the software provided met the needs of the client group in a variety of different respects, leading to a reflective implementation so that users would be able to adapt the use of the software in the light of their own practices and needs.

There are a variety of evaluation models (House, 1978), however as the most important issue was the fitting of the software to the client needs, a needs-based evaluation was felt to be the most appropriate model.

Focus groups were held with faculty representatives from each of the 3 faculties at Heriot-Watt University. These were individuals who had previous experience of designing assessments. Two focus groups were held using the same two facilitators at each using a predetermined schedule. This schedule was designed to get the participants thinking of the wider assessment needs that they have without being constrained by the desire to appear practical, thus avoiding some of the difficulties associated with a narrow focus on accessible needs (Judd et al., 1991), such as the concentration on measurable results, by discussing in general terms the context of assessment and the future directions that they saw assessment heading.

The results of the focus groups were written up by one of the facilitators and approved by the other and key issues and themes were drawn into the needs of the participants. Technical members of staff were also consulted about the practical issues of designing a web-based assessment engine and their responses were added to the needs and requirements of a web-based assessment engine which informed the development and design of the WebTest engine (Crofts and Tomes, 1997).

Background

Heriot Watt University has a number of particular features many of which informed the development of the WebTest assessment engine: It is primarily a campus university based on a greenfield site at Riccarton on the outskirts of Edinburgh; there is a very high male to female ratio, and a strong science/engineering/management bias. Heriot Watt University has also been

in the forefront of learning technology developments for a number of years, and thus has the infrastructure to cope with new innovations that might well be lacking in some more traditional institutions. These factors obviously informed the development of the WebTest engine, through the requirements that people demanded from the system within the focus groups. Although it is clear that such a bias exists, trying to second guess the consequences of that would be a minefield and would go entirely against the development principles of the WebTest engine that informed its development in the first place.

A great deal of useful information was gathered from the needs analysis which can be brought to bear on any future evaluation. Obrecht (1999) suggests that the most effective way to generate the needs of the users is to use an expert panel with high credibility within and outside the organisation that is at arm's length to both the organisation and the external stakeholders. The evaluative methodology that the SCAAN project is implementing to match users requirements with an appropriate web-based assessment engine involves using the needs demanded by the development groups at Heriot-Watt University to supplement any issues which decision makers unilaterally generate.

The idea behind such an evaluative framework is to ensure that the needs of the users are at the forefront of the mind when decisions are being taken about which assessment engine to implement. Certain assessment engines have particular strengths in certain areas while others are better at other things. WebTest for example is extremely good at displaying mathematical notation - a product of having being developed in such a scientific University, but has no authoring interface - questions must be entered in XML; Miranda has a very friendly authoring interface, but the question types are limited; TRIADs has a wide range of question styles that the engine can implement, however it requires plug-in which can cause installation problems for the novice computer user wishing to connect from home. Allowing people the space to think critically about what they actually require from a web based assessment engine means that they are more able to make informed choices on a wide range of criteria.

This methodology also makes explicit the stakeholders involved in the choice of the assessment engine - meaning that there is less chance that certain users are sidelined and their concerns subsumed by the wishes of the budget-holder.

There is very little information around on which to make an informed choice about web engine. Most decisions are currently made on the basis of price and of connection to the people developing the software - with the major commercial product - QuestionMark's Perception seen as the benchmark or default choice - although certain institutions have become involved with the development, or have adopted other engines.

The role of the SCAAN consortium is to raise awareness in Scottish HE of the availability of web based computer assisted - which involves raising awareness that there are genuine choices to be made between engines - rather than simply plumping for the popular commercial product - not only is this expensive, but it may well be less well suited to their needs than one of the other products which is available - but until users start to analyse their

needs this will remain hidden. This evaluation framework gives the novice CAA user the ability to make an informed choice of CAA product, taking into account the needs of all of the stakeholders in the system.

The Evaluative Framework

The evaluation comes in two parts - designed each to be completed within an afternoon. The requirements analysis examines what the users feel that they require from the engine that they adopt; while the satisfaction analysis examines how well the piece of software that they have chosen meets their actual needs in a practical context.

Requirements Analysis

First of all the requirements analysis looks at what the users feel that they require from the engine that they adopt. Before using an assessment engine each group of users (students; lecturers; systems administrators and central administrators) would participate in a thirty minute brainstorming session followed by completion of the Criteria Importance Quantification schedule for their user group (see Appendix) which is based on the results of the focus groups held at Heriot Watt. This methodology allows participants the space to reflect on their needs in a CAA system without imposition from the experience of Heriot Watt, while retaining the practical advantages of tapping into the issues that Heriot Watt has identified through extensive consultation. The brainstorming session is designed to highlight any issues which are particular to that department/institution and also to draw up any unidentified general criteria which have not been thought to be of importance by Heriot Watt stakeholders. These issues should be added on to the end of the Criteria Importance Quantification schedule (CIQ; see Appendix). The CIQ exercise recognised that different departments and institutions have different priorities and needs, and aims to assess their relative importance. The eventual aim of the exercise is to generate a profile of user group needs, which can then be compared to the different engines.

There is a danger that when using a Likart Scale as part of a needs assessment, users may feel obliged to score everything as highly desirable in order to maximise the power of the system. This can cause problems as such a user will have low discrimination between the items and thus their relative weight will be diminished. This should be explained to participants, and they should be encouraged to utilise the full scale. Where this has been a significant factor, difficulties of interpretation and communication can arise as the mean scores may be skewed. This skew can be corrected by rescaling each participant's scores to a mean of 3. Although this does not correct the low weight of the participant, it does aid the interpretation of the results.

Standardisation Procedure

1. Total the item scores of each participant (Pt)
2. Divide Pt by the number of items on the schedule for that user group (Pm)
3. Subtract 3 from Pm to give the standardisation amount
4. Subtract the standardisation amount from the item scores

Example

A user has provided data from a 5-item schedule of 3,4,4,5,3.
Pt=19
Pm=3.8
Standardisation amount = 0.8
Rescaled scores become 2.2, 3.2, 3.2, 4.2, 2.2

Totalling the participant scores across each of the items can draw up a profile of departmental needs. A high score indicates that the participants have felt that this is a high priority need for the assessment engine adopted to address.

There are further analyses that can be performed to further inform the choice of engine.

High standard deviation within an item indicates that not all members of the user group have similar priorities within an area. Where there are a large number of items of high standard deviation (<1.98) it should be ensured that all of the participants understand the items. If so, it may well be that organisational and management factors are concealing user subgroups that may have different priorities. This may well be the case in multidisciplinary areas such as Latin American Studies, where the linguists may well feel that their priorities lie with an engine which has a great deal of support for multimedia capabilities, while the literature specialists may desire an engine which incorporates the assessment of free text. In such a case it is desirable to re-evaluate whether it is appropriate to adopt one engine for the department, or whether more than one engine should be adopted - perhaps looking outwith the department for support.

Weightings can be attached to the priorities of the different user groups, dependent on the resources available in a department - a department which has a high flexibility on system support issues, for example, can deprioritise the needs of the system analysts on the grounds that additional support can be drawn on if necessary.

Once a profile of the department's priorities has been drawn up this can then be compared to the profile of the available engines. The SCAAN consortium has gone some way to draw together the features of the different engines which they are involved with (SCAAN Consortium, 1999), although it is hoped that more detailed profiles can be drawn together by the end of the project.

Satisfaction Analysis

After the department has implemented their eventual choice of engine. A satisfaction analysis can be conducted to see how well the engine adopted met the needs of users.

Again each user group would be isolated and allowed about an hour for the completion of the analysis. This would take the form of a Criteria Satisfaction Quantification schedule (CSQ), with each of the items identical to those of the CIQ, however the primary question has changed to ask how well the engine met their needs on each of the criteria points; followed by a period of general reflective discussion.

The CSQ exercise allows data to be gathered on how satisfied they were with the tool with respect to criteria that they have identified as being important to them. The personal reflection time allows the opportunity to bring up any issues that they felt they had not adequately considered before the trial, but which with hindsight have proved to be of importance.

It should be explained to participants that in the CSQ, the engine should be assessed on how well it met their actual needs, rather than theoretical needs. If for example, the engine adopted was not web compliant, but the participant had no need of internet capabilities then a "3" should be recorded; if this was a positive advantage (perhaps because of security issues for example) then a "4" or "5" would be an appropriate response. A "1" or "2" should not be recorded simply because the engine did not have internet capabilities if these were not a desirable attribute in the eyes of the participant.

The scores recorded on the CSQ can be compared to the results of the CIQ and the profile of the engine adopted to assess whether the trial was successful, and whether use of the engine should be maintained or another trial undertaken with another engine.

In particular, consideration should be given to any additional requirements that are highlighted in the reflective exercise to discover whether there are any requirements that are difficult to identify without experiencing the software. Where all required needs were not fulfilled for practical or resourcing reasons, review can be undertaken to assess whether these deficiencies are noted in the CSQs and/or in the reflective period, and whether consideration should be given to the adoption of an engine which would satisfy the user demands better.

Discussion

This evaluative framework provides a quantitative means of evaluating the needs of the users of a CAA system in a small-scale setting informed by qualitative data, taking into consideration the views of the stakeholders in the implementation.

The framework is informed by the work of Heriot Watt University, in developing a set of user criteria. This has some drawbacks in that members of the user group may be swayed by items which appear on the schedule but which are irrelevant in their own institution, however the advantages of drawing on an extensive piece of qualitative research, using participants well versed in assessment issues, which would be impractical to carry out on a small scale.

The process itself does involve a time commitment on the part of users. Although University staff are under considerable pressure, research has

shown that stress is most often a product of a feeling of lack of control over the working environment.

Where there are more than one engine which appears to meet the needs of the department, and consideration is being given to adoption of an alternative trial should the first prove unsuccessful, it would be wise to consider whether any content generated in the engine (such as questions) would be portable, so that if a change was deemed necessary, no work would be lost. At the time of writing, it would be difficult to transport content generated in one engine to another, however with the introduction of IMS standards (http://www.imsproject.org/xml) and the announced intention of the prominent commercial firm QuestionMark to allow exportation of Perception content in an IMS compatible format, this might not be such an issue in future. Within the SCAAN consortium, comparison of the Miranda and WebTest document type definition has been undertaken (SCAAN, 2000) and it is hoped to produce software which would export both Miranda and WebTest content into an IMS compatible form. This would allow questions produced in Miranda, WebTest and Perception each to be read by all of the engines. Although this may solve the problem of redundant content, transferral between one engine to another can still prove a time consuming exercise, most engines require some familiarity with a particular format, both at the time of question generation and for students using the engine, and it must be remembered that certain features may not be supported in an alternative format, even if the core question is not lost.

One of the current problems of assessing which engine is most appropriate for adoption within a department is the lack of information about the availability of assessment engines and the features of them. The SCAAN consortium is currently undertaking dissemination of the TRIADs, Miranda and WebTest assessment engines and comparison of the engines has already been partially completed. It is hoped that the results generated by the evaluation of the dissemination trial can further inform this comparison and allow a direct comparison of user requirements and the features of available engines.

Uptake of CAA in Scotland is not ubiquitous although some form is used at all HE establishments. The difficulties of choosing an appropriate engine and the consequences of a wrong engine being implemented in terms of time, money and effort are delaying the more widespread uptake of this technology and reaping its potential. It is hoped that this framework can provide a means of adopting a suitable, efficient engine which will serve the needs of the user groups providing small-scale potential adopters with a means of confidently selecting an engine and encouraging them to participate in the transformation of assessment.

References

Beishuizen (1999) Study Strategies in a Computer Assisted Study Environment Learning and Instruction; v9 n3

Bennington (1999) Ethical Implications of Computer-Mediated Evaluation. New Directions for Evaluation; n84

Crofts (1999) Enabling Reuse of CAA by Design , 3rd International CAA Conference, Loughborough

Foster and Crofts (1997) WebTest and Mathematics Service Courses, Maths & Stats Newsletter, November 1997

House (1978) Assumptions Underlying Evaluation Models Educational Researcher V7

IMS Project (15/5/00, 13.53 GMT) IMS XML Bindings, DTDs and Examples (http://imsproject.org/xml/)

Judd, Smith and Kidder (1991) Research Methods in Social Relations, Holt, Rinehart and Wilson, New York.

Mochrie (1998) A Diet of Carrots: Autonomy in Learning Mathematics for Economics, in Mogey (ed.) Evaluation Studies, LTDI, Edinburgh

Russell (1999) Testing on Computers: A Follow-up Study Comparing Performance on Computer and on Paper. Education Policy Analysis Archives; v7 n20

Sclater and Howie (2000) The SCAAN Project, 4th International CAA Conference, Loughborough

SCAAN Consortium (1999) Technical Issues Report, SHEFC, Edinburgh

SCAAN Consortium (2000) DTD Comparison Document for the SCAAN Project, SHEFC, Edinburgh

Criteria Importance Schedule: Administrators

Instructions

The following criteria of a good web-based assessment system from the administrators' point of view were developed at Heriot Watt University following extensive consultation. Please examine how important each criterion is from your perspective and indicate by circling 1-5 where 1 is "of no importance" and 5 is "critical".

In your opinion, how important is it that any web-based assessment systems should…

...be compatible with the administrative information system in use?

Data should be held in a form that can be easily input to any central administrative administration system.

Of no importance critical

1 2 3 4 5

...provide an audit trail for tests contributing to final assessment?

For significant tests auditing should be implemented to allow any queries in results to be answered.

Of no importance critical

1 2 3 4 5

...fulfil any statutory requirements?

The system should provide the University with the tools to fulfil any statutory requirements associated with assessment.

Of no importance critical

1 2 3 4 5

...be future-proof by using open standards?

By using open standards no particular product is favoured and existing software from commercial, shareware and public domain may be used within the system as well as the possible future products which work to these standards.

Of no importance critical

1 2 3 4 5

...use "worldware"?

"worldware" are those (software) products which do generic tasks such as WWW browsers, email clients, word processors, spreadsheets etc. Using "worldware" taps into the vast investment in these products and allows students/lecturers/developers to use existing skills

Of no importance critical

1 2 3 4 5

...separate content from the assessment engine?

By separating content from the assessment engine, the content is available for other purposes and the engine can be upgraded independent of content.

Of no importance critical

1 2 3 4 5

...store content in re-useable form?

The major investment in a CAA is the content. This material contains intellectual property which should not be wasted and should be kept available for future/other uses.

Of no importance critical

1 2 3 4 5

...be platform independent in delivery?

Delivery may be anywhere on campus, but also on other sites of the University and to distance learning students. The system should be deliverable on any reasonably current platform (hardware and OS)

Of no importance critical

1 2 3 4 5

...be robust?

Students make unpredictable responses to questions. No response should cause the system to fail.

Of no importance critical

1 2 3 4 5

...be technically reliable?

The system may be used at any time and for tests of varying significance including some contributing to formal assessment. The system must therefore be highly technically reliable.

Of no importance				critical
1	**2**	**3**	**4**	**5**

...be scalable, particularly in delivery?

The system must be able to scale in size as needs increase.

Of no importance				critical
1	**2**	**3**	**4**	**5**

...be extendible in functionality?

New subjects may require new types of answers and new methods of judging these answers. Advances in other software may provide new modes for entering answers such as 3D models in chemistry, geology, architecture etc. The system should be extendible to include these future demands/ opportunities.

Of no importance				critical
1	**2**	**3**	**4**	**5**

...require minimal support in delivery?

The system should provide tools to minimise the support required in day-to-day delivery of the tests. The resources required for supporting large numbers of students and tests should be proportional to the benefit to the University.

Of no importance				critical
1	**2**	**3**	**4**	**5**

...use generic/transferable skills in all stages?

The core skills required to develop and run the system must be generic/transferable. Staff can then be drawn from/used in other areas/applications in the University.

Of no importance				critical
1	**2**	**3**	**4**	**5**

...provide a simple, efficient system for content capture and quality assurance?

The system should minimise the resources required to develop tests without compromising on standards of delivery and quality. The system should not be a "barrier" to development of tests.

Of no importance				critical
1	**2**	**3**	**4**	**5**

...be compatible with existing information systems?

The content of the tests and the data (results, logs etc) generated by the system should be in a form compatible with existing information systems

Of no importance				critical
1	**2**	**3**	**4**	**5**

...provide a simple efficient interface for lecturers to access their students' data?

The system should allow lecturers easy access to the information they require from it.

Of no importance				critical
1	**2**	**3**	**4**	**5**

...be maintainable in the long term?

The system should be practical to maintain for several years.

Of no importance				critical
1	**2**	**3**	**4**	**5**

Criteria Importance Schedule: Lecturers

Instructions

The following criteria of a good web-based assessment system from the lecturers' point of view were developed at Heriot Watt University following extensive consultation. Please examine how important each criterion is from your perspective and indicate by circling 1-5 where 1 is "of no importance" and 5 is " critical".

In your opinion, how important is it that any web-based assessment systems should…

…set the questions lecturers want to ask ?

Lecturers should be able to adopt whichever teaching style they think appropriate.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…require no more effort from lecturer than equivalent paper tests

Lecturers should be able to define the questions in an easy, familiar mode and have them implemented within an acceptable time-scale.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…give lecturers control over the delivery of tests

The system should allow lecturers to control the delivery of tests eg time window for taking the test, no of attempts, questions delivered all at once or sequentially, their use as self, formative or summative assessment.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…display questions to accepted standards of typesetting and presentation

Most subjects have standards for presenting specialist data. These should be adhered to as far as possible.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…judge answers appropriately

Lecturers should understand the algorithms used to judge answers and agree their robustness.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…set marking system required by lecturer

The system should allow any reasonable marking system to be applied.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…provide secure system for managing marks

The system should store the marks in a form accessible only to those with appropriate rights.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…provide reports on student(s)

The system should allow a lecturer to review progress of individual students and groups of students

Of no importance				critical
1	**2**	**3**	**4**	**5**

…allow feedback lecturer to/from students

The system should allow two-way feedback between lecturer and student.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…provide reports on questions

The system should provide reports on particular questions to highlight any question which may create problems for the students - through misunderstanding or faulty definition

Of no importance				critical
1	**2**	**3**	**4**	**5**

…provide an audit trail if requested

The system should provide an audit trail which allows a lecturer to respond to students querying results.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…keep content in a re-useable form

The system should keep the test and question definitions in a format that can be use with other systems and other purposes such as paper documents.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…allow integration of/with other learning resources

It should be possible to integrate other learning resources or provide links to them.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…establish the IPR of the content

The system should display origins of content where appropriate.

Of no importance				critical
1	**2**	**3**	**4**	**5**

Criteria Importance Schedule: Students

Instructions

The following criteria of a good web-based assessment system from the students' point of view were developed at Heriot Watt University following extensive consultation. Please examine how important each criterion is from your perspective and indicate by circling 1-5 where 1 is "of no importance" and 5 is " critical".

In your opinion, how important is it that any web-based assessment systems should…

…test knowledge of the subject not the system?

The user interface should require the minimum of training and should allow students to answer questions as naturally as possible.

Of no importance critical

1 2 3 4 5

…provide reliable answer-judging?

The algorithms used to judge answers should reliably detect right and wrong answers.

Of no importance critical

1 2 3 4 5

…provide an opportunity to comment on tests?

Students should have the opportunity to communicate problems and other comments directly to the lecturer. This facility should be integrated with the CAA system so that the students can make their comments as they take the test and the lecturer can refer to the specific test being taken if it is relevant.

Of no importance critical

1 2 3 4 5

…keep students marks secure and confidential?

An appropriate level of security and confidentiality should be provided.

Of no importance critical

1 2 3 4 5

…give appropriate feedback to student?

The feedback should be relevant to the question, and given in a time-scale appropriate to the type of test.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…be physically accessible?

The system should be available across the campus(es) and in distance-learning mode.

Of no importance				critical
1	**2**	**3**	**4**	**5**

…be accessible for special needs?

No student should be disadvantaged by CAA and any special needs should be either addressed by the system or alternative assessment provided.

Of no importance				critical
1	**2**	**3**	**4**	**5**