SQA Guidelines on Online Assessment for Further Education

Contents

Foreword

SQA has been exploring issues of online assessment for several years. In February 2001 we updated the SQA guidance on open and distance learning and began to address some online assessment and quality assurance issues. Since then, the Pass-IT project has begun — a significant inter-agency research project. It will pilot online assessment of SQA qualifications and generate definitive guidance by October 2004.

We are aware that in Further Education SQA qualifications are already being adapted for online assessment, and that the sector requires guidance from SQA sooner rather than later. For this reason the Scottish Further Education Funding Council (SFEFC) agreed to support the production of these interim *SQA Guidelines on Online Assessment for Further Education.*

The project was initiated in summer 2002 and was directed by a steering group of representatives from the Further Education sector, including colleges which participated in the SFEFC-funded *Best Practice in Online Assessment* project. A consultant was commissioned to produce the guidance.

The guidelines address Computer Assisted Assessment (CAA) and online issues and offer technical advice on:

♦   why you might choose to use online assessment — and when not to use it
♦   what you will need to do to use online assessment — including staff development and technical support
♦   software and hardware issues
♦   developing online assessments — deciding on question types
♦   ensuring the quality of the items
♦   deciding on the delivery methods
♦   security and authentication features, including invigilated test sessions
♦   supporting candidates and designing feedback

We are about to begin a second SFEFC-funded project which will examine e-moderation issues and produce guidance for centres and for SQA Moderators by summer 2003. We are collaborating with Learning and Teaching Scotland, the Scottish Further Education Unit, the Scottish Centre for Research into Online Learning and Assessment (SCROLLA) and BBC Scotland to research online assessment issues during the Pass-IT project. In addition, we are working with other UK examining and awarding bodies to explore the quality assurance and technical issues related to online assessment.

The intention is that these research initiatives will inform the development of future SQA guidance on online assessment. For this reason the *Guidelines on Online Assessment for Further Education* should be viewed as a work in progress — and as a technical support document which should be read in conjunction with the existing SQA publication *Assessment and Quality Assurance for Open and*

Distance Learning (February 2001, A1030). Both of these publications will be updated and merged with the guidance to be produced at the end of the Pass-IT project in October 2004.

In issuing this publication at this time we hope to facilitate the use of online assessment in Further Education centres, and would ask you to work with SQA to raise and address the assessment and quality assurance issues which will inform the longer term guidance on online assessment.

Introduction and overview

Introduction

This document provides guidance for those in Scottish Further Education colleges who are using computers and internet technologies to assess candidates for SQA qualifications. These guidelines focus on online summative assessment within centres, though you may find them useful when planning online formative assessment. There is also guidance on carrying out remote online assessment.

The guidelines are divided into four main parts: *analysing your requirements*, *preparing to assess online*, *holding invigilated assessment sessions* and *post-assessment issues*. Each of these represents a stage in implementing online assessment, and for each stage we recommend a series of concrete steps to take. The order these steps are presented in is not prescriptive, and requirements will differ between centres, some of which have already gained considerable experience of online assessment.

While these guidelines are aimed particularly at those who wish to implement online assessment, much of the content is also relevant in other situations such as classroom-based and remote assessing using pen and paper.

Overview: what is online assessment?

Computers have been used to assess candidates for decades. Software to allow assessments to be delivered to candidates (often referred to as computer-assisted assessment, or CAA), became widely commercially available in the 1990s, and many institutions began to experiment with these packages. The increasing availability of networked computers by the mid-1990s allowed assessments and other educational services to be delivered online with web browsers. Centres no longer had to think about installing CAA software on candidates' machines and manually gathering results from these machines after assessments had taken place. Instead they could put the CAA system on a central server and rely on a standard web browser to deliver assessments to candidates and send responses back to the server for storage, automatic marking and return of feedback to the candidates.

In these guidelines the term *online assessment* involves the use of internet-based tools for:

♦ the creation, storage and delivery of assessments to candidates
♦ the capture, marking, storage and analysis of their responses
♦ the collation, return and analysis of results

Scope of these guidelines

Online assessment is used for three main purposes:

♦ *diagnostic* — to ascertain the level of a learner's (or group's) knowledge, carried out before or after learning activities take place

♦ *formative* — carried out during a learning programme after a learning activity has been carried out to determine its effectiveness

♦ *summative* — carried out at the end of the learning programme or at a number of specific points during it to provide a mark which affects the level of qualification achieved or allows progression to the next stage of learning

Summative assessment is generally broader in scope than formative assessment, which tends to focus on a smaller part of the curriculum. It also normally examines the retention of information and the integration of skills over a longer period of time. These guidelines refer to the processes surrounding online summative assessment, though much of the information will be relevant to diagnostic or formative assessment. It is, in any case, advisable to trial all systems extensively in a formative mode before carrying out summative assessments.

We will look specifically at delivering assessment using web browsers, though much of the document is relevant to other CAA technologies such as offline delivery and the use of personal digital assistants (PDAs).

The document is aimed primarily at centre management. However, various sections are also highly relevant to lecturers, staff developers and technical support staff.

♦ If you are involved in formulating your centre's strategy for the use of online assessment you should read section 1: *Analysing your requirements*.

♦ Section 2: *Developing online assessments* should be read by anyone who is creating items and assessments — or selecting them from an item bank.

♦ Those involved in the summative testing process should read sections 3: *Holding invigilated assessment sessions* and 4: *Post-assessment issues.*

1 Analysing your requirements

Before beginning to develop questions and deliver them to candidates, you should analyse your reasons for delivering online assessment, the SQA Units to be targeted initially, and whether your centre has the necessary organisational and technical infrastructure. Your reasons should be explained to staff who will be involved in the online assessment process. Candidates should also be told why online assessment is being introduced, and how it links with other course activities.

1.1 Ensure institutional commitment

a High level support

Because of the range of skills and technologies required in order to deliver online assessment, it is essential that your centre's management gives its support — both financial and moral. You should prepare a clear strategy for the implementation of online assessments, outlining the target SQA Units and the implications for staff, and disseminate it across the centre.

b Costs

Many centres are now using virtual learning environments (VLEs) with basic online assessment capabilities. There are a few commercial online assessment packages which offer considerably better assessment functionality than is found in these VLEs. Some are not cheap and may have additional hardware and support requirements. However, the costs of backing up data, hardware maintenance, systems administration, training of technical staff, and integration with other centre information systems might already be covered if you are using your VLE. There would also be no additional administrative overhead in registering candidates with the system and handing out usernames and passwords.

Other costs which need to be considered whether or not you use your VLE for assessing are:

♦ **Obtaining or creating questions and assessments.** Creating objective assessments is more time-consuming than, for example, developing essay questions.

♦ **Staff development** to assist staff in producing questions, using the systems, supporting candidates and invigilating.

♦ **Technical support** for staff and candidates, particularly when using the system for the first time and during online examinations.

Your centre might be able to offset such additional costs against reductions in marking time and administration. Economies of scale are increasingly likely as the uptake of online assessment increases across the centre.

c   Staff development

While centre managers may find a briefing on online assessment adequate, teaching staff who are deploying the technologies are likely to need training and support. Support staff in libraries, and in administration and technical areas, may need specialised training in how to support those using the systems.

♦ **Management:** senior centre staff and heads of departments should be advised of the reasons for implementing online assessment, the scope of the initial trials and likely roll out, and the implications for other services.

♦ **Teaching staff:** the vendor of the product you select may be able to offer training. Other organisations such as Netskills (www.netskills.ac.uk) may also have relevant training courses. Many courses concentrate on how to use the tools and largely ignore pedagogical issues. Parts of this document — in particular section 2 — should be of assistance to staff tasked with creating assessments and items. It may be advisable for your staff developers to put together a training programme aimed specifically at meeting the needs of centre staff.

♦ **Support staff:** VLE vendors and some of the larger assessment system vendors may offer training for systems administrators. Other support staff, such as librarians, are unlikely to find appropriate courses specifically aimed at them, but will need to be thoroughly familiar with the system and the way the centre intends to use it. They should be encouraged to attend external or in-house training.

d   Technical support

Technical support will be required at a number of levels. Maintenance of computer labs, networks and servers becomes much more critical when an invigilated assessment is taking place using this equipment. It is advisable that:

♦ technical staff are given advanced notice and made aware that such assessments should be given priority

♦ at least two staff at the centre know how to administer the assessment software on the server

♦ before and during assessments dedicated technical support is available

e   Copyright

There should be a named person within each institution with responsibility for advising on intellectual property rights and copyright clearance. Section 2.4 provides more information on checking copyright.

1.2 Define your reasons for using online assessment

There are a number of good reasons for assessing candidates online. Centres should be clear which benefits they are likely to realise before embarking on online assessment.

a   **Improving learning**

One of the main reasons for introducing online assessment is the positive impact it can make on the learning process. Developing objective items can lead to better quality questions, while using assessment software to analyse questions can help you to improve them.

Objective online assessments can be delivered throughout the delivery of the Unit and give instant feedback to candidates. Increasing the frequency of assessment and improving feedback are recognised motivators for learners, and they can also help staff to identify weaker candidates who need remedial assistance. Many candidates will only do coursework which is assessed, so frequent online assessments can ensure that they learn more without increasing marking loads.

Formative activities can reinforce the learning while preparing candidates for summative assessments using the same system.

Finally, online assessment allows media-rich assessments which include video, sound, simulations and animations. These can be used to provide more realistic scenarios and case studies, and are particularly useful for subject areas such as languages.

b   **Objective marking**

Marking can vary between markers; computers, however, will mark the same responses in the same way every time. The downside of this is that current assessment software can be less accommodating than human markers — for example, where a candidate is entering text, a computer might mark something wrong which a person would consider correct. However, if you stick to questions with a strictly limited range of responses you can ensure that assessments will be marked accurately and consistently. See paragraph g 'When not to use online assessment' overleaf for other situations where objective marking may not be appropriate.

c   **Reducing marking load**

One of the most commonly cited reasons for introducing online assessment is to reduce the burden of marking. With objective online assessments all the marking can be automated — the benefits for assessors and institutions are obvious. Don't forget, though, that there will be a need for increased human input in areas such as question authoring, training, and technical support.

d   Automated results collation, analysis and return

Another major benefit of online assessment is the ability to collate results and instantly analyse the performance of a single candidate, a class, or a number of classes. Some assessment packages have sophisticated analysis tools with features such as graphing facilities and calculation of means. Questions and assessments can also be analysed and made more reliable and valid over time.

Once of the most significant administrative benefits is the automatic transfer of results into management information systems, which saves time and reduces clerical errors. Marks can also be collated easily for awarding bodies, and can be provided to candidates online.

There will be a reduction in costs for the printing of examination papers and the associated administration involved in distributing papers and collecting responses. With robust online systems, the chances of a candidate's responses going missing should be reduced.

e   Improving access

Candidates with disabilities may find online assessment preferable to traditional forms involving pencil and paper. Because the assessments can be rendered automatically in ways which suit the candidate, it should be possible to increase the text size or have the questions read out by a screen reader. For candidates confined to the home or living in remote locations, the ability to take an assessment via the internet may be helpful.

f   Reducing opportunities for plagiarism

Essays submitted electronically can be analysed using various technologies to check for plagiarism. Inconsistencies between the style of different sections, suggesting incorporation of others' work, can be detected automatically. Software is also available to locate similar materials on the internet and to compare the work of a cohort of candidates (see section 4.2 for further information on detecting plagiarism).

g   When not to use online assessment

Objective questions, such as multiple choice questions (MCQs), may not always be appropriate. They are effective in assessing factual recall, but it is harder to use them to assess higher order learning skills such as creativity, communication skills or synthesis. They cannot elicit qualitative or reflective responses. It is also clear that candidates' understanding and problem-solving skills are assessed better when they have to recall and enter information than when they are presented with a selection of possible responses.

Online methods such as short essays, asynchronous discussions and chat, where there is a high level of human input, can be used in almost any assessment situation if there are clear advantages (such as for distance education). For obvious reasons, though, it may not be possible or advisable to assess practical

skills (such as the use of a lathe) entirely using online assessment. You should use your experience and professional judgement to determine the most appropriate assessment tool.

SQA does not recommend using objective assessments exclusively — they should be combined with other methods such as essays and practical tasks. The validity and reliability of the overall assessment should therefore be enhanced by combining several forms of assessment. The appropriateness of the assessment tool — not reducing the administrative burden — should be the primary driver for introducing online assessment.

1.3 Decide on delivery method

a Invigilated assessment sessions

Section 3 details the processes surrounding invigilated assessment sessions. These allow a high degree of authentication of candidates, in the same way that traditional paper-based examinations do. You also have more control over the materials available to candidates during the assessment, and over who they can communicate with.

b Remote assessment

In open and distance learning situations it is often impractical to bring candidates to a single location for assessment so other strategies must be developed. The internet provides an opportunity for giving candidates in remote locations the same assessment at the same time. Using online assessment means that the candidate's submission is transferred to the centre immediately, and the candidate can be advised of their achievement for any objective questions instantly. SQA's *Assessment and Quality Assurance for Open and Distance Learning* (February 2001, A1030) covers the issues you will need to consider if assessing remotely.

Authentication

Despite the many ways in which the internet can assist in remote assessing, authentication remains a serious problem. There are three types of authentication available:

♦ *knowledge* — something only the candidate should *know*, such as password or a PIN number

♦ *ownership* — an identifier which the candidate must *own* such as a smart card

♦ *characteristic* — a unique *characteristic* of the candidate such as their iris or fingerprint

Currently only knowledge IDs tend to be used for online assessment due to the cost of the other technologies. This also remains the case for other critical applications, such as online banking and shopping.

Impersonation without the candidate's permission

It is possible that someone might want to use a candidate's authentication details without their approval. A combination of authentication methods such as knowledge and ownership (as when you take money from an ATM) would help to avoid this, but is not yet practical or affordable. With an assessment available at a single specified time, a candidate should instantly be aware if someone else has authenticated using their details because they will not be able to log into the assessment system. In this case the centre should be able to take immediate action to issue a new password.

Assessments can also sometimes be restricted to specific computers or groups of computers (based on their IP addresses or domains). This is helpful where candidates are accessing machines with permanent connections to the internet, but less so when temporarily logging in through an Internet Service Provider (ISP) as the IP address cannot be determined in advance. Restricting delivery of the assessment to a particular ISP is possible, however.

Impersonation with the candidate's permission

If the candidate wishes someone else to complete the assessment on their behalf, they could hand over passwords and smart cards with ease. You can employ various strategies though to help ensure that candidates are who they say they are:

♦ **Use real time questioning.** Online assessments delivered regularly throughout a Unit can be backed up later by telephone or face-to-face interviews with candidates to assess their level of knowledge.

♦ **Get to know writing style.** Assessors will often get to know a candidate's style of writing. Ensuring that summative tasks incorporate an element of writing, and comparing this with writing delivered throughout the Unit, can help.

Restricting access to other resources

It is impossible to ensure that candidates do not have access to books or even another computer with internet access while carrying out an unsupervised assessment remotely. With timed assessments they may not have much opportunity to access other resources anyway, particularly where such assessments assess understanding rather than knowledge. Only open-book assessments are feasible in such situations. It may help if you require candidates to declare evidence of any resources they used to complete the assessment.

Practicalities of home-based assessment

Candidates should, of course, be fully familiar with their operating systems and the assessment software before sitting any timed summative assessment. Because you have no control over their environment, you cannot ensure that their equipment is functional or that no one else is using their telephone when they try to link to their ISP. Their system may also be running slowly due, for instance, to a full hard disk or a slow processor. Summative assessments which run on a basic web browser and do not require additional technologies such as Java-enabled browsers or plug-ins are more likely to be delivered successfully. Candidates

should have already taken formative assessments using identical technologies to those required for the summative session.

You should also be aware that bandwidth is likely to be lower for home users, so multimedia materials may be impractical — and also be aware that the candidate may be paying for access to the internet. Systems which allow formative assessments to be taken in offline mode as well as online will help to reduce internet access costs.

Logging, electronic certificates and reminders

Your assessment system should maintain a record of when and from what IP address assessments were submitted. Ideally, candidates will be sent an electronic certificate through e-mail with a unique identifier which records the date and time of submission. Later, it should be possible for candidates to receive results and feedback through e-mail.

Most assessments for distance candidates will not be required to be completed at specific times but will have a deadline. When candidates submit an assessment after the due date, it may be helpful if the system prompts them for a reason for the late submission and provides an appropriate warning. The warning might, for instance, inform a candidate that they should contact their tutor. It may also be possible to set up the system so that a candidate is prompted automatically by e-mail shortly before an assessment is due or shortly afterwards if the assessment has not been submitted.

Invigilated remote assessing

Many of the difficulties of remote assessing can be solved if the assessment is supervised by a responsible person. It may be possible for candidates to use a local assessment centre where the environment can be controlled, or to take the assessment through their employer who can also authenticate them. You should be satisfied, though, that these arrangements do not compromise the confidentiality of the assessment.

c   Closed v open-book assessments

The issues surrounding whether to allow candidates to bring materials of their choice to invigilated online assessments are similar to those in invigilated paper-based assessments. In addition, if allowing open-book assessments, you will need to ensure there is adequate space at the candidate's workstation to place books and other materials.

You may also wish to allow access to software tools such as calculators and spell-checkers, internet-based resources, and even communication facilities such as chat. It can be useful to assess the candidate's skills in finding resources for an assessment on the internet. Network demands can be considerably higher if this is permitted — your network administrator should be aware of whether this will be feasible. In practice, the time restrictions of an invigilated session can mean that there is little opportunity for tracking down resources or collaboration.

d   Electronic portfolios

Electronic or digital portfolios are being used increasingly to record and structure candidates' learning activities, and can be used for assessment. They can demonstrate skills in the collection, organisation and interpretation of a variety of documents and sources of information. They can store hyperlinks, text documents, graphics, video and other media, study diaries, and records of practical work — and all of these can be used in assessing candidates' work and competence. It is now possible to integrate electronic portfolio software closely with VLEs and assessment systems.

1.4   Select appropriate software

a   Is your VLE adequate?

As noted in section 1.1b, you may be able to carry out online assessment using your existing VLE. However, VLE assessment facilities can be inferior to those of software packages whose prime function is assessment. Recognising this, some of the leading VLE vendors now allow particular assessment systems to be linked to their systems to enable common authentication and return of results for storage in the VLE. You should examine the functionality of your VLE and compare it with that of some of the leading online assessment systems to see if it would meet your online assessment requirements.

b   Quality of vendor

As with the selection of any software which is being rolled out across the institution, you should choose a vendor with a good reputation whose products are recommended by users elsewhere. Good support for the product is essential, as are regular upgrades which incorporate bug fixes and new features. Products with a large number of users tend to have national or international user support networks and discussion lists, which can be helpful.

c   Navigation facilities and prompts

Navigation between items should be straightforward and consistent. It should be possible to ignore questions and return to them at a later stage, or to alter the response to a previously answered question. A facility to allow candidates to go to any item by number, or to the first or last item, is also helpful. If candidates are permitted to leave the assessment early, they should be prompted for confirmation that they have completed the assessment. It is normally possible for candidates to ignore items which they cannot answer, though in certain circumstances it should be possible to require them to input a response before the next item is presented. Some systems also enable candidates to flag a question that they would later like to return to.

d   Usability

Candidates should have some control over how the assessment is presented. In particular, they must be able to increase text size or use a screen reader. Buttons and icons should be positioned and named consistently. Messages should appear in a consistent format and be easy to understand.

e   Help facilities

Candidates should have had sufficient practice in using the system before taking an online summative assessment. However, a help facility should be available throughout the assessment which explains how to respond to items, navigate through them, and quit. In addition, information about the numbers of items which have been answered and unanswered should be displayed, together with the time remaining for the assessment. It should be possible to have a warning message sent automatically to the candidate ten minutes before the end of the assessment.

f   Security features

There are many measures which can be taken to improve security during online assessments. Some possible security breaches which can be prevented with the help of the software are:

♦ **Loss of candidates' responses:** there should be measures to prevent accidental quitting, such as a dialogue box requesting confirmation. In the event of system failure or network outage, responses should have been saved so that candidates can start from where they left off when the system has been restored. Timing should be reset accordingly.

♦ **Communication with others and access to information sources:** for summative assessment, some systems have their own browser which does not permit printing or access to other parts of the internet. Assessment data should be able to be encrypted to reduce the possibilities of its interception between server and browser. Disabling or removing other communication software, such as chat or e-mail, should also be considered.

♦ **Minimising opportunities for copying:** it should be possible to offer candidates different versions of the same assessment. The software should be able to select items randomly from an item bank, and to shuffle the order of items and the order in which options are presented. Different parameters in mathematical questions should be able to be delivered to different candidates. There is, however, evidence to show that the selection and ordering of items and options can affect performance, so randomisation such as this should be used with care. It may be fairer to use the software to deliver at random one of a number of pre-structured, and moderated, assessments, than to present candidates with an assessment with a high degree of randomisation created at runtime.

g   Candidates' responses

The system should ensure that items and responses are not cached on the candidate's machine.

It should be possible for you to view details of when a candidate took an assessment, how long it took, and from which IP address it was taken. You should be able to re-create the assessment exactly as it was delivered to the candidate with any randomly generated parameters, items or ordering. Records should be kept of responses to each item, including the full text of open-ended responses and any feedback or result provided to the candidate.

h   Interoperability

Online assessment systems can be used in isolation from other systems, but the administrative benefits would be minimal. Ideally, you need to be able to import items and assessments created by people using other systems, to be able to export your assessments to other systems, and to be able to export responses and results to spreadsheets, statistical analysis packages and candidate record databases.

The only internationally-recognised specification for the transfer of assessment data is the IMS Question and Test Interoperability (QTI) specification. It uses EXtensible Markup Language (XML), which has become the world-wide standard for storing and exchanging structured data over the internet. QTI can be used in conjunction with other XML standards, such as MathML, which incorporates mathematical and scientific notation.

The QTI specification is in two parts: *Assessment, Section and Item (ASI)* deals with the content of the assessment; while *Results Reporting* specifies how results should be exchanged. Each can be implemented independently. A number of vendors of VLEs and assessment systems now claim that their systems allow the import and export of assessment data in QTI format. Most work only at the item level, so any data relating to assessments will be lost when exporting.

It is advisable to assess vendors' claims of interoperability rigorously before purchasing a system. One reason for this is that there are various ways to interpret parts of the QTI specification. The Centre for Educational Technology Interoperability Standards (CETIS) Assessment Special Interest Group (www.cetis.ac.uk/assessment), funded by JISC, brings together authors of a number of assessment systems and is developing a common interpretation of the specification to reduce its ambiguities. The Technologies for Online Interoperable Assessment (TOIA) project (www.toia.ac.uk), also funded by JISC, is currently developing an assessment system for UK further and higher education which implements this common interpretation of QTI.

i   Accessibility

By law you should take all reasonable steps not to discriminate against candidates with disabilities. You should make a judgement as to whether online assessment is appropriate for candidates with special needs, and whether a non-IT alternative

can be provided. In any case, you must ensure that item validity is not compromised.

Assistive technologies that can be deployed include interfaces to screen readers, speech recognition software and touch screens. Make sure you allocate more time for candidates if the use of such devices requires extra time. If sound input or output is required, candidates may have to be provided with headphones or be located in a separate room. Dyslexic candidates should have their text response submissions reviewed manually unless correct word order and spelling are learning requirements.

You should always ensure that images are provided with a textual equivalent, and that candidates with poor eyesight can adjust the text size (which should not be hard-coded). Some assessment systems allow the use of style sheets, enabling users to customise the appearance of web pages.

Some item types are unsuitable for candidates with particular disabilities. For example, drag and drop type items and hotspot items may not be appropriate for the visually impaired or those with poor motor skills.

Good design for disabled candidates is often good design for all candidates. For instance, the use of large radio buttons for multiple-choice questions may also assist nervous candidates having trouble using a mouse.

TechDIS (www.techdis.ac.uk), a JISC service, provides further information on how to assist candidates with special needs.

j   Other software considerations

You will also need to check whether there are any variations in the ways assessments are rendered on different platforms and with different web browsers and versions. Some online assessment packages require the use of plug-ins such as Flash which, though widely available, will still need to be installed on every candidate's machine. Others require the use of Java applets or JavaScript; the ability to render assessments with these technologies varies widely between browsers, and sometimes may require to be enabled in the browser first. In addition, you may have to ensure that specific fonts are installed for foreign languages.

On the server side, you should, of course, ensure that the assessment system runs on the operating system of your server, or that you can purchase suitable equipment. You should also have the systems administration skills available to install and maintain software on the recommended platform. Some of the leading assessment systems are not, for example, available on UNIX or LINUX platforms, which may be the platform of choice for other software in your centre.

In addition, any database system used by the assessment software for storing results should be compatible with the systems you already have in place.

1.5 Develop hardware infrastructure

a Servers

Servers should be on maintenance contracts with their vendors so that immediate repairs can be carried out in the event of failure. It is advisable to mirror your server in a separate location so that assessments can still go ahead if the primary server breaks down.

Assessment software vendors should be able to provide details of the predicted load on hardware and the server side resources required per user of the software, as well as the minimum hardware requirements.

Servers should be backed up daily, and it is particularly important to back up candidate submissions immediately after assessments.

b Network infrastructure

By specifying the likely frequency and size of traffic between browser and server during one assessment, and the number of candidates, it should be possible to predict the load on the network. Don't forget to consider other uses of the network which may be taking place at the same time as the assessment. If network capacity could be a potential problem, you could eliminate any unnecessary multimedia in the assessments.

You could allow for the duration of assessments to be extended in the event of any network delays, but it is much better to ensure that network capacity is adequate in the first place. It is also advisable to adapt your appeals procedures to incorporate a strategy for dealing with network and other technical problems. These procedures should be approved by SQA.

c Candidate PCs

Online assessment system vendors will specify the minimum requirements for the memory and speed of candidate PCs, and their operating system and version. You should also check the screen resolution and colour depth, and the keyboards of all machines which will be used by candidates.

The language of the operating system determines, for example, how dates and numbers are displayed. It is particularly important to make sure that there is no opportunity for such settings to be tampered with in advance of an assessment. No candidate should be put at any disadvantage by the specification or set up of the equipment they are using.

1.6 Select target Units

You are advised to start with a pilot before rolling out online assessment extensively. Target Units should be identified. Lecturers responsible for these Units will need to have a reasonable degree of computer literacy, a willingness to

help deliver (and possibly develop) online assessments for the Units, and an ability to enthuse candidates about the new method of assessment.

Target Units should have an appropriate number of candidates — with a small number of candidates the benefits of online assessment will be difficult to demonstrate. On the other hand, too many candidates might require the upgrading of computing facilities and unfeasibly high levels of support.

You should then decide which parts of the Units could be assessed with the help of computers. How often should assessments be carried out? Should taking them be a requirement for passing, or should the mark for each one count towards the final mark for the Unit?

When developing online assessments for existing Units you should use your experience in relation to the assessments that are being replaced. If you are concerned that creating or adapting an assessment instrument for the online environment will alter the skills and knowledge to be assessed by the Unit, you should contact SQA.

2 Developing online assessments

'Online assessment' can refer to the online submission of candidate evidence and the return of feedback and results to them. This can be done using an assessment system or simply by using e-mail.

It can also mean the delivery of assessment to be completed and marked online. In this case, an assessment consists of a number of 'items'. The BS7988: *Code of Practice in the use of information technology (IT) in the delivery of assessments* sets out procedures which are required to ensure that online assessments are delivered professionally. BS7988 defines an item as the smallest *separately identified question or task within an assessment*. However, an item normally consists of more than just the question and can include data such as the responses, the correct response, the score and feedback for each response. (For more on BS7988, see section 3 of this guide.)

2.1 Choose appropriate item types

Most people start off with multiple-choice questions (MCQs) when developing online assessments. There is no need, though, to restrict yourself to MCQs, as there is a growing range of item types available. Most of these are objective item types where the range of answers is highly restricted and can therefore be marked automatically. Essay-marking software is under development, and is already proving effective in marking short essays (see section 2.7).

The choice of item type should be appropriate to the knowledge or skill being assessed. Bloom's taxonomy of questioning specifies six levels of competence which can be assessed: knowledge, comprehension, application, analysis, synthesis and evaluation. MCQs tend to be used to assess lower-level competences such as knowledge and comprehension. However, with skill and imagination it may be possible to use objective items to assess some of the other competences.

The main item types you are likely to find in an online assessment package are:

Multiple choice

We're all familiar with MCQs, where candidates are required to choose one out of several responses to a question.

Name the capital of Saudi Arabia.
- ○ Mecca
- ○ Medina
- ○ Jiddah
- ○ Riyadh

True/false and yes/no

True/false is a simple type of MCQ with two possible responses. Yes/no questions are almost identical.

| Sydney is the capital of Australia. | Is Sydney the capital of Australia? |
| ○ True   ○ False | ○ Yes   ○ No |

Such questions allow the rapid assessment of large amounts of material. However, a score of 50% is statistically likely in an assessment containing only true/false or yes/no items if the candidate knows nothing about the subject. Also, it can be difficult to write stems which have two such absolute responses.

Multiple response

Similar to MCQs but much harder to answer correctly are multiple response questions where the candidate can choose one or more correct responses.

Which of the following islands are found in the Firth of Clyde?
☐ Arran
☐ Bute
☐ Cumbrae
☐ Gigha

Notice how the convention is to use a square *checkbox* rather than a round *radio button* where more than one response can be selected. This is a function of the software, and is not something you can normally alter.

Matching

This question type can be used when candidates are required to match two related items or concepts. It may require the candidate to drag the related items beside each other.

Associate the following plants with the most appropriate type:

lysichiton americanus	evergreen perennial
cornus alba	deciduous shrub
phormium tenax	marginal aquatic perennial
verbascum bombyciferum	biennial or short-lived evergreen perennial

Matching questions are variations of MCQs. In the example above, for the first plant you must choose one of four types, the second plant one of three and the third plant one of two. This means that if the candidate does not know what type the last plant is they can still get the question right by a process of elimination. It may therefore be a good idea in this example to incorporate an extra plant type (ie an extra category on the right) — both to provide a more taxing question, and to ensure that the candidate is not forced into a second error if they have already selected one incorrect option.

Sequencing

Sometimes candidates are required to know a sequence. They will normally be required to drag the responses into the correct order.

Put the following colours of the rainbow into the correct order.

Green
Orange
Violet
Yellow
Indigo
Blue
Red

Hotspot

Hotspot questions are useful when a candidate should to be able to identify or interpret parts of a picture or diagram.

Click the stigma of this flower.
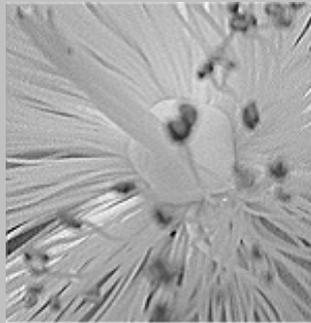
Short answer

You may wish to present the user with a question or problem which requires one of a small, finite number of textual responses which you have pre-defined.

A body will continue in its state of rest or uniform velocity unless acted upon by an unbalanced force.

What is this law known as?

In this case you would have to define all the acceptable permutations of the correct response, eg *Newton's First Law of Motion*, *newton's first law of motion*. A problem with this kind of question is that you may fail to specify a response such as *Isaac Newton's first law of motion* that would not normally be marked wrong by a human marker of a physics assessment. Some assessment systems can be set up to search for a particular word or sequence of words within a response, which may alleviate this issue.

In the next example there is only one correct answer (and no necessity for the candidate to enter mathematical notation).

A 200m high waterfall has water flowing through it at a rate of 100kg per second. What is the rise in temperature of the water between the top and bottom of the waterfall?

☐ °C

Fill in the blank/gap fill

You can present candidates with a phrase, sentence or paragraph which is missing one or more words. They are then required to complete the text either by typing the missing word or selecting it from a pull-down list. The first example here is a variation of the short answer question type, the second a kind of multiple choice.

Complete the following:

Now is the ☐ of our discontent.

Toledo ☐ es / está ☐ a setenta kilómetros de Madrid.

Essay

While automated essay marking software is in its infancy, you may still wish candidates to enter short essay style answers which can be marked by a person later. To give greater depth to some of the above objective question types, for example, you could ask candidates to justify their responses using a short essay style question.

Toledo ☐ es / está ☐ a setenta kilómetros de Madrid.

Justify your choice of es or está in the above sentence.

☐

2.2 Create valid items

a Ensuring quality in all aspects of the item

A valid item should enable you to assess whether a learning objective has been met. Items consist of various parts, all of which need to be valid. A multiple choice item, for instance, is composed of a question or statement (*stem*) and a number of possible responses (*options*). The correct option is known as the *key*, the incorrect ones known as *distractors*. For all aspects of the item:

♦ Appropriate grammar and orthodox spelling are essential.
♦ Word and sentence order should be straightforward.

- Use negatives only when strictly necessary, and embolden them to reduce ambiguity. Double negatives should be avoided completely.
- Avoid the use of technical terms or words which are not directly related to the skills and knowledge the item is assessing.
- Ensure technical soundness (see section 2.5).
- Ensure that the wording is equally appropriate to all well-prepared candidates whatever their age, gender or ethnicity.
- Carefully consider the amount of text and graphics presented to the candidate. It is harder to read from the screen than from paper, and the use of scrolling can add time and difficulty to the assessment.

b Creating stems

- Stems should normally contain a single unambiguous statement or question.
- They should be concise and clear, and should contain only information which is directly relevant.
- It should be possible for an able candidate to come up with the correct response when given the stem only.
- Avoid duplicating material in each option — include it in the stem.
- Avoid giving clues to the key.
- Present information in the best logical order. If you include a graphic, label it at the top to avoid the candidate wasting time trying to work out what it means.

c Creating keys and distractors

A distractor should aim to present the weaker candidates with plausible alternatives.

- Any factual errors in a distractor may be noticed by the candidate who could then rule it out and be more likely to guess the correct answer.
- Distractors should be plausible.
- Options should be shorter than stems to avoid excessive reading.
- The options following a stem should be of a similar length, style and type to reduce the possibility of guessing.
- Ensure that there is only one key.
- Ensure that you have actually included a key!
- 'All of the above' should be avoided. A candidate who identifies an incorrect distractor would then also be able to rule out an 'all of the above' distractor if there is one.
- Using 'None of the above' is also not a good idea. If you make it the key you are still unable to discover if the candidate has the requisite knowledge.
- Don't use absolutes such as 'always' and 'never' as, except in mathematics, such words may be hard to prove.

♦ Avoid qualifications such as 'sometimes' and 'usually' — options with these words tend to be true.

♦ Ensure that distractors are significantly different from each other so that candidates who know the correct response are not confused.

♦ Three or four options are normally sufficient. It can be hard to think of any more plausible distractors for each item.

d   Adding complexity to MCQs

MCQs are often criticised for being used to assess basic factual knowledge rather than deep learning. However, a case study or problem which requires to be understood, analysed or worked out, followed by a stem and a number of plausible distractors, can be used to assess higher level learning. Such items take longer to answer, so should have more marks allocated to them than simpler ones. You should also avoid candidates having to scroll up and down through large amounts of case material on the screen.

One way to assess the ability to solve more complex problems is to build a multiple completion MCQ item such as the following:

> An inelastic collision takes place between a moving object and a stationary object. Both objects have the same mass. In this situation, which of the following quantities is/are conserved for this system?
>
> 1 Momentum
> 2 Total energy
> 3 Kinetic energy
>
> ○ 3
> ○ 1 and 2
> ○ 2 and 3
> ○ 1, 2 and 3

2.3   Add metadata

Some online assessment systems allow you to add descriptive information (known as metadata) to an item or assessment so that it can later be searched for and retrieved. As item banks are built up, it becomes increasingly important to use metadata. With funding from SFEFC, the Colleges Open Learning Group (COLEG) is currently developing a large item bank for Scottish FE — each item will incorporate various metadata fields to facilitate searching.

There is as yet no internationally agreed schema for assessment-specific metadata. However, it is possible to select a subset of metadata fields from initiatives such as the Institute of Electrical and Electronic Engineers' (IEEE) Learning Object Metadata (LOM). For example, the Electronics and Electrical Engineering Assessment Network (www.e3an.ac.uk) has chosen the following metadata fields to be included with all its items:

1. Type of item (multiple choice/multiple response/numeric etc).
2. Time — expected to take in minutes.
3. Level — (introductory/intermediate etc — in Scottish FE it would be appropriate to use SCQF levels).
4. Discrimination — (threshold students/good students/excellent students).
5. Cognitive level — (knowledge/understanding/application/analysis/ synthesis/evaluation).
6. Style — (formative/summative/formative or summative/diagnostic).
7. Theme — subject.
8. Subthemes — what part of that subject.
9. Related themes — what other themes might find this question useful.
10. Description — text for use by people browsing the database.
11. Keywords — for use by people searching the database.
12. Validation — confirmation that the question has been validated (moderated).
13. Peer review date.
14. Reviewer(s) — who reviewed it.
15. Comments — made by the reviewer.

The CETIS Metadata Special Interest Group (www.cetis.ac.uk/metadata) provides further information on metadata.

2.4   Check copyright

If all items used in your assessments have been created internally, you are unlikely to be infringing copyright. However, you should carefully check your use of all text, graphics and other elements in the items.

If an item, or any part of it, was created elsewhere (eg downloaded from the internet), you must ensure that written permission has been given to you for its use for educational purposes. Exceptions to this may include websites from which public domain resources are freely downloadable. Even with these, however, you should take care, as the authors of the website may not have cleared copyright themselves for the materials they are making available.

If you are using items from an item bank managed by another party, the other party should ensure that copyright clearance for all items has been obtained and be able to provide you with a policy statement outlining its approach to copyright clearance.

Colleges should identify a person within the institution who is responsible for advising on intellectual property rights and copyright clearance. The JISC Legal Information Service (www.jlis.ac.uk) can provide you with further guidance.

2.5   Analyse item quality

There are various measures which can help you analyse the quality of an item. A major advantage of carrying out assessment online is that some of the software packages can assist you in measuring the quality of items and help you to decide whether they should be retained, rejected or enhanced in some way.

Facility is a measure of the difficulty of an item. It ranges from 0 to 1. A score of 0 implies an impossibly difficult question, while an item with a facility of 1 is too easy. A facility of 0.5 means that 50% of candidates get the question right.

The *discrimination* index helps you discover how well the item distinguishes between candidates of different abilities. The value will range from –1 to +1. Discrimination of greater than +0.2 should be adequate. If it turns out that less able candidates are getting a question right while able candidates are choosing the wrong response, then the item will have a negative discrimination index and is not appropriate.

The number of times an option is selected can also be measured. This is sometimes referred to as *frequency* analysis. If a distractor is rarely chosen, it should be replaced.

If objective items are to be used in summative assessments, they should be tested beforehand with an appropriate group of students.

2.6   Create valid and reliable assessments

When you have created a number of items, you will need to combine them into an assessment. (You can also create an assessment by selecting items from an item bank.) The composition of the assessment should cover all the learning outcomes. If there is any variation in the assessment presented to each candidate, you must ensure that it is of an equivalent standard and level of difficulty.

The number of items in an assessment should be determined by the total time available and the likely length of time required to complete each item. The BS7988 standard (see section 3) recommends that candidates should spend no longer than 1.5 hours sitting at the computer. This means that an assessment should be broken into separate sessions if it would take longer than this to complete. However, SQA has conducted a Review of National Qualifications and it has been agreed that a Unit assessment involving the assessment of knowledge and understanding should take no longer that 1 hour, or 1.5 hours when assessing both practical skills and knowledge and understanding.

The score for an item is meaningless outside the context of an assessment, so this information is usually included at assessment rather than item level. See section 2.8 for further information on scoring strategies.

You should ensure not only that individual assessments are reliable but also that these will lead to a reliable final assessment decision over the Unit as a whole.

2.7 Incorporate automated text marking

Items require either the *selection* or the *generation* of a response. MCQs and most objective item types involve selection from a series of options, but items that require the generation of a response can be a better assessment of both recall and understanding.

Where text is generated there are obvious administrative advantages in being able to have it marked automatically — assuming it can be done as validly and reliably as it would be by a human marker. The two main approaches are *keyword analysis* and *natural language processing*.

Keyword analysis looks for key words or phrases in the text to see if candidates have included these in their response. The system will need to be able to accept alternative spellings or mis-spellings, synonyms, punctuation or lack of it, variations in upper and lower case and units for numerical answers. As existing systems may accept invalid responses or fail to accept valid responses, all responses of this type should be checked by a human marker when they are being used for summative purposes.

Natural language processing looks for much more than keywords, and is already being used to mark sentences and compensate for different word order, synonyms and misspelling. Ultimately, a balance has to be struck between the cost-effectiveness, speed and consistency of computer-marked assessments and the professional judgement of the human marker who can ensure that at least partial credit is given for poorly-expressed responses.

A system has been developed by the US-based Educational Testing Service (ETS) to analyse full essays and produce scores for organisation, sentence structure and content. Essays which have already been graded by human markers based on these characteristics are fed into the system enabling it to estimate the likely score a reader would give to an essay submitted on the same topic.

Ultimately we will see sophisticated text marking systems such as these incorporated in online assessment software, but for the moment the available systems have limited capabilities. If using such systems, a provisional mark may be awarded automatically, but for the time being it will also be necessary for a human assessor to assess essay-type responses.

2.8 Design a scoring strategy

One attribute of an assessment which needs to be specified is its scoring strategy. Candidates should be made aware of how items are scored within an assessment and how the overall result is calculated.

a Cut-off scores

Deciding on a cut-off score which distinguishes between acceptable and unacceptable performance in an assessment is a complex issue. A cut-off score of

50% may be advisable where background knowledge and understanding are being assessed. However, 100% may be appropriate where critical knowledge (eg landing a plane) is being assessed. Where the assessor wishes to ascertain mastery of the subject, a cut-off score of 70–80% may be applied. There may also be a requirement to discriminate between degrees of acceptable performance by allocating, for example, A, B and C grades.

Any valid distinctions the original assessment instrument made between those who are found competent and those who are judged not (yet) competent should be carried forward into the new online instrument as far as possible.

b   Scoring MCQs

With MCQs, the usual policy is to allocate one mark for a correct response and none for an item which is ignored or incorrectly answered. A particularly difficult item can have a larger number of marks allocated to it.

Negative marking can be used to attempt to reduce the effects of guessing. One strategy is to allocate a mark for each correct answer, subtract a mark for each incorrect answer, and allocate no marks for an unanswered item. Some people feel that negative marking is unnecessary if assessments have been well constructed.

c   Scoring multiple response items

When scoring multiple response items there are at least two possible scoring strategies:

♦   If you answer the item entirely correctly you receive full marks, otherwise you receive nothing.

♦   Every correct response receives a mark while every incorrect one means a mark is deducted.

2.9   Design feedback method

There are various options for giving feedback to candidates on their performance. These will depend on the nature and purpose of the assessment. Scores or grades can be provided for a Unit overall, a particular assessment, a section within the assessment, or an individual item. Feedback relating to correct responses can also be given with reasons, explanations, links to learning resources, and hints for further study. However, in summative assessment, it is important that feedback does not compromise the confidentiality of items which may be used with future candidates.

2.10 Design for re-use

You can develop an online assessment and deliver it just once to a group of candidates. CAA becomes more viable, however, if your assessments can be re-used and delivered in different ways and contexts.

♦ **Exchanging questions.** Some institutions already pool their expertise in developing questions. One example in higher education is the Electronics and Electrical Engineering Assessment Network (www.e3an.ac.uk). It makes sense not to replicate the effort of producing high quality questions in every centre but to share content. It is also possible to buy in assessments from commercial content developers.

Being able to share assessment content such as questions, assessments and results with others requires them to be put into formats which are common to everyone's software systems (see section 1.4h 'Interoperability'). However, many CAA packages and VLEs do not currently allow you to export this data out of their own proprietary format. So, unless everyone is using the same CAA package or VLE, it could be difficult to work together. Fortunately, Scottish colleges now have VLEs so it will at least be possible to exchange assessment content within an institution.

♦ **Future-proofing your assessments.** You will want to ensure that the time spent keying in questions is not wasted if you or your institution upgrades to a different system later on. To future-proof your work, you should ensure that assessment content can be output in a non-proprietary format which is recognised by various assessment systems — the only such format currently available is the IMS Question and Test Interoperability specification (see section 1.4h Interoperability).

♦ **Delivering on different platforms.** You may want to provide facilities for your candidates to take assessments on a variety of platforms, such as Personal Digital Assistants (PDAs) and web browsers. You could also use different assessment systems for different purposes. You might, for instance, deliver questions via a virtual learning environment (VLE) for formative purposes, then later make some changes and export them to a high-specification secure assessment system for delivering formal examinations.
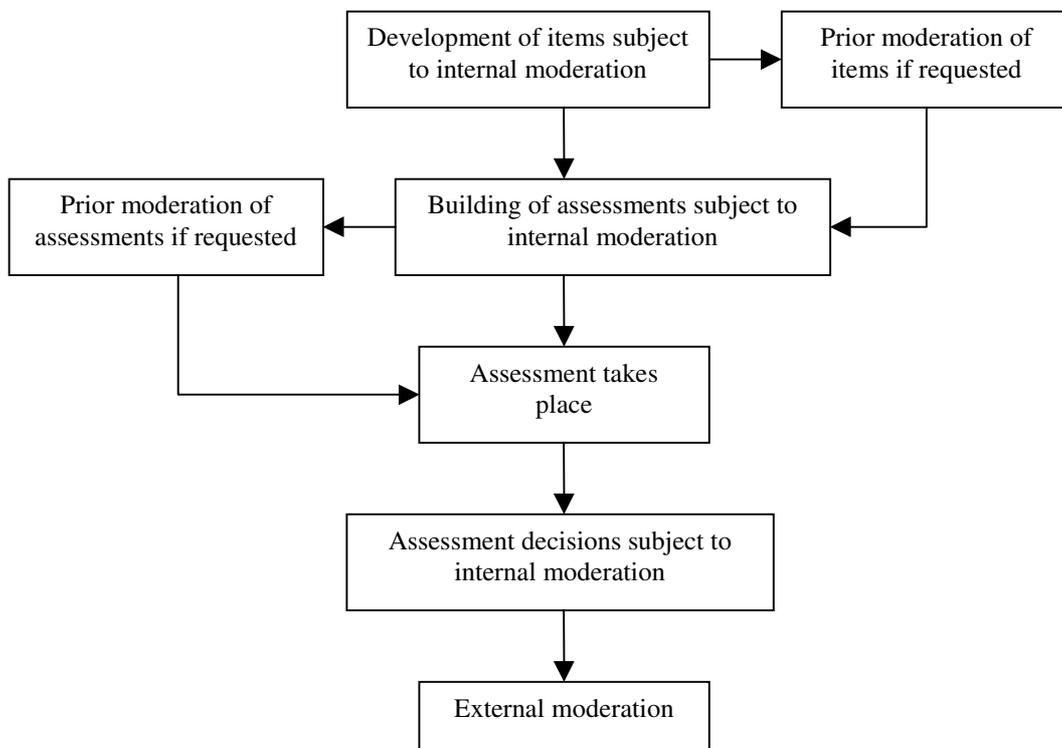
2.11 Use item banks

You may wish to use an item bank to provide a different, but equivalent, set of items for each candidate. You should be able to demonstrate that these items are indeed equivalent and that the bank is big enough to provide different versions of the assessment. It should also be able to provide an entirely new set of questions to candidates if they are being re-assessed.

2.12 Moderate assessments

Moderation procedures are set out in the SQA publications:

♦ *Quality Assurance Principles, Elements and Criteria*
   (December 1998, A0798)
♦ *Assessment and Quality Assurance for Open and Distance Learning*
   (February 2001, A1030)
♦ *Guide to Assessment and Quality Assurance for Colleges of Further
   Education* (December 2001, A0841/2)

Moderation activities for online assessments are the same as for other forms of
assessment (see the diagram), unless the assessment has been developed from a
prior-moderated item bank with reference to a pre-determined test specification.

```
┌──────────────────────┐          ┌──────────────────────┐
│ Development of items  │─────────▶│  Prior moderation of │
│ subject to internal   │          │  items if requested  │
│ moderation            │          │                      │
└──────────────────────┘          └──────────────────────┘
            │                                  │
            ▼                                  ▼
┌────────────────┐   ┌──────────────────────────────┐
│ Prior moderation│◀──│ Building of assessments       │◀──
│ of assessments  │   │ subject to internal moderation│
│ if requested    │   └──────────────────────────────┘
└────────────────┘                  │
        │                           ▼
        │               ┌──────────────────────┐
        └──────────────▶│  Assessment takes     │
                        │  place                │
                        └──────────────────────┘
                                    │
                                    ▼
                        ┌──────────────────────┐
                        │ Assessment decisions  │
                        │ subject to internal   │
                        │ moderation            │
                        └──────────────────────┘
                                    │
                                    ▼
                        ┌──────────────────────┐
                        │  External moderation  │
                        └──────────────────────┘
```

3 Holding invigilated assessment sessions

When conditions for assessment require assessments to be supervised, you should take steps to ensure that they are delivered in an environment that both supports candidates and upholds the conditions for assessment.

BS7988: C*ode of Practice in the use of information technology (IT) in the delivery of assessments* sets out procedures which are required to ensure that online assessments are delivered professionally. SQA does not require centres to comply with this standard, but BS7988 does offer valuable advice, and you may wish to consider complying with the relevant parts. This section of these guidelines has been written with the BS7988 standard in mind.

3.1 Develop evaluation procedures

Consider how you are going to evaluate use of the system. You may wish to observe candidates, invigilators and support staff using the systems, to interview them and to give them questionnaires on their experiences. The impact on the network, and on the availability of computer labs for other purposes could also be evaluated.

3.2 Trial system

Before you use the system with candidates, it should be fully checked with all hardware and software platforms on which assessments may be delivered. This should include:

♦ registration and authentication facilities
♦ the provision of the appropriate assessment at the right time and within any time limits
♦ the correct delivery (and randomisation) of items, parameters and response options
♦ the correct delivery of feedback for options
♦ correct navigation through the items
♦ the recording of responses

3.3 Register candidates for system

Candidates should be given individual usernames and passwords for the system. If these are the same as their e-mail or other user accounts, administration will be minimal.

3.4   Give candidates practice

Candidates should be able to practice using the system for formative assessments well in advance of the summative assessment. These practice assessments should include navigational aspects of the software, together with all item types which will be used later — items should have a similar style and difficulty. Candidates should be given feedback and results in the same way as they can expect from the summative assessment.

Candidates should be encouraged not to work on the computer for long periods and to take the opportunity to look away from the screen and relax their muscles periodically.

Feedback should be sought from candidates during formative assessments on all aspects of the hardware, software, items and operational procedures so that these can be refined in advance of the summative assessment.

3.5   Give candidates prior information

Candidates should be advised in advance of:

♦   the dates and duration of assessments

♦   authentication procedures and any ID required

♦   the knowledge which may be assessed and any subdivisions of the assessment

♦   the number and types of items in the assessment, and how these will be scored

♦   any permitted, non-permitted or provided data or aids, such as calculators (physical or software)

♦   provision for candidates with special needs

♦   procedures for dealing with irregularities such as cheating

♦   appeals procedure including what to do in the event of technical problems

♦   how scores are used and to whom these are communicated

3.6   Prepare room and equipment

Well in advance of an assessment you will need to assess the room in which it will take place and its equipment.

♦   The room itself should have suitable temperature, ventilation and lighting for working both on and off screen. There should be no glare on the screens.

♦   Candidates taking the same assessments should not be able to see the screens of others; there should be partitions or a minimum distance of 1.25m between workstations.

♦   You may also need to provide additional space for candidates in wheelchairs or those requiring assistive technologies. If any such technologies could be

distracting to other candidates, an additional room and workstations should be provided.

♦ Seats should be adjustable and the height of the workstation and screen should be correct.

♦ There should be space for printed materials and other equipment, and space for making rough notes if required.

♦ If there is any sound output with the assessment, headphones must be provided.

♦ The invigilator's desk should be in a good position to spot communication with others and the use of unauthorised reference material, and it should be possible for the invigilator to walk around the room viewing every workstation.

♦ Virus protection should be in place and up-to-date on candidates' PCs.

Immediately before the assessment takes place, a number of checks should be made on the equipment:

♦ All PCs and monitors should be switched on to ensure that they are working.

♦ Keyboards and mice should be functioning.

♦ The browser should be present, correctly configured, and able to access the assessment system.

♦ Access to unauthorised software should be disabled.

3.7   Ensure adequate staffing

More staff will be required in the process than in paper-based exams due to the necessity to have technical support available at immediate notice. Network administrators and technical support staff must be advised of the time for the assessment in advance, and must treat it as a priority.

Those carrying out supervisory or invigilation functions will need to be familiar with the assessment system, how to log in, navigate, respond to all item types and quit, though it should not be necessary to assist candidates with these functions if adequate opportunities to practice have been provided. They should not be distracted from invigilation duties by providing technical assistance, so there may be a need for more than one invigilator. Invigilators should also be careful to assist only in the use of the system and not in understanding the content of any items. There may be specific functions available to invigilators, such as being able to reset timers after a break due to system failure. Other requirements for invigilators are to:

♦ Manage the identification of candidates, complete records of attendance and check that these correspond with the number of candidates present.

♦ Start and end the assessment.

♦ Reduce opportunities for collusion and the use of non-permitted aids and materials by patrolling the room.

♦ Ensure that candidates do not remove materials from the room except where permitted.

♦ Ensure that final submit buttons have been pressed and that all web browsers are closed after candidates have left the room.

3.8 Ensure security

Measures need to be taken to ensure that assessments and items are available only to authenticated candidates at the permitted times. Candidates should have no opportunity to access unauthorised materials or to communicate with or copy from others unless this is permitted for the assessment. One way to detect unauthorised activity is to install software which captures all keystrokes. In addition to the recommended security features of the assessment software, which are outlined in section 1.4f 'Security features', the following procedures are recommended.

♦ **Avoiding unauthorised access to the system.** The assessment software should be on a secure server with password protection at different levels to protect items, especially indications of the correct responses. The identity of the candidate should be verified with an ID card unless known to the invigilator. Assessments can be restricted to a particular domain or set of IP addresses, and it may also be advisable to allocate individual machines to users.

♦ **Removing access to unauthorised materials.** Aids such as calculators and spell checkers and access to unauthorised sources of information should be disabled as necessary in advance. This can be achieved by hiding navigation toolbars, disabling shortcut keys and preventing return to the assessment if another program has been accessed. You should also prevent the candidate from viewing the source code of the web page in the browser.

♦ **Preventing unauthorised disclosure of content.** Facilities to print, copy materials into another application, or send files to the hard disk, removable media and other computers should be removed. There should be no item or response data left in caches or in the file system after the assessment has taken place.

♦ **Be alert for the possibility of cheating or malpractice.** Invigilators should log incidences where the conditions for assessment outlined for the activity are breached or malpractice is suspected.

3.9 Specify technical failure procedures

Invigilators must be able to reset timers after an unplanned break. In this case candidates, who are only permitted to access the assessment a single time, must be able to go back to it if authorised by the invigilator.

It is advisable that appropriately qualified staff are on hand to bring up the backup server if the main one goes down. Technical support should also be at the ready with spare PCs, monitors, keyboards and mice in case this equipment fails. It is advisable to have one or two spare workstations if this happens — the invigilator must be able to authorise candidates to restart the assessment on another PC without any loss of working time, and to access any responses they have already made.

The invigilator should log all technical failures and delays, and the complaints of any candidates which could lead to an appeal. No candidate should be disadvantaged by technical problems.

Finally, if there is any doubt about the proper functioning of any aspect of the hardware, software or operations (particularly when carrying out initial pilots), you may wish to have paper copies of the assessment to hand!

3.10 Specify emergency procedures

In the event of fire alarms or other emergencies, invigilators must be able to lock the room securely with browsers still open. Access to server rooms must also be safeguarded at such times. Candidates should have minimal opportunities for collusion during such interruptions. When they return to the assessment they must be given the full time for completion. Details of all emergencies and the implications for the assessment must be logged by the invigilator.

3.11 Authenticate candidates

Candidates should already have practised using the system with their authentication details. In reality, one or more of them may have forgotten their username and/or password, so it should be possible to re-advise them of these if necessary.

The assessment should not begin until the candidate has had a chance to read the initial instructions and information about the number of items, the item types being used, the scoring mechanism and the duration of the assessment.

4 Post-assessment issues

4.1 Mark assessments

If any human input is required for marking an assessment you may want to ensure this is done anonymously, in which case you should ensure that no details about the candidate are provided to the marker.

4.2 Check for plagiarism

Candidates should be made aware of centre policies on plagiarism and the resulting disciplinary procedures if they are found to be claiming someone else's work as their own. One reason why plagiarism may be on the increase in Scottish colleges is that subject materials can be obtained electronically and easily inserted into a submission. As reported in the media, there are websites which contain banks of essays for purchase. However these are often of poor quality and their style is likely to differ from the normal writing style of the candidate. Under exam conditions plagiarism can be largely eradicated, but there are also several technologies which can help to detect it when candidates are submitting written work at other times:

- Linguistic analysis can be carried out on the submission to detect changes in its grammatical style thus identifying sections which may have been copied from elsewhere. This helps to detect material from textbooks.
- Submissions from different candidates and those of previous years can be compared to detect similarities.
- The internet can be searched for similar or identical work. One problem is that no search tool can cover the whole internet. Also you may be required to pay to access a particular essay!

Creating a climate where plagiarism is discouraged is ultimately the best policy. Opportunities for copying can be reduced if new assignments are set each year. Creative and analytical tasks such as contrasting two approaches are less likely to have been carried out before, and this will lessen opportunities for collusion. It may also be useful to assess the process as well as the product — each candidate should be able to demonstrate an individual approach to how they produced their assessment. A further suggestion is that assessments could be submitted in stages, such as an initial plan, a draft and a final version, allowing you to see how they have changed.

Carrol and Appleton (2001) (see Further reading, page 34) detail many other helpful suggestions for reducing possibilities of plagiarism, how to detect it and what action to take if it occurs. Bull et al (2001) have analysed some of the software available to help detect plagiarism. JISC has also set up a national plagiarism advisory service at www.jisc.ac.uk/plagiarism.

4.3   Ensure secure storage of assessment data

There should be no unauthorised access to assessment data such as assessments, items, candidate submissions and achievements. The Data Protection Act must be followed with regard to personal data. All such data should be retained by the centre in case of queries and appeals. It should also be held for purposes of moderation. The SQA *Guide to Assessment and Quality Assurance for Colleges of Further Education* (March 2003, A0841/3) provides information on how long to store evidence for moderation purposes. Data should be backed up regularly with backups kept in secure locations.

4.4   Return results to candidates

Online assessment systems allow you to return results to candidates. This can be done instantly in the case of objective assessments, though it is advisable to provide access to results only after the centre has checked that everything has worked correctly. Bear in mind that a large number of candidates checking their results simultaneously can have a serious impact on the network and the functioning of the assessment system.

4.5   Deal with appeals

If a candidate makes an appeal against an online internal assessment you should ensure that your assessment software allows you to recreate the assessment, incorporating any randomisation exactly as the candidate has been presented with it. The normal college appeals procedure should then be followed.

4.6   Return results to SQA

The dispatch and receipt of results to SQA should be recorded in the usual way through the college internal system, and as defined in the Operational Guide for the current year.

4.7   Analyse results

Various facilities are available in most online assessment systems to analyse the results of candidates. You may wish to calculate class means, compare results with other cohorts, or add up scores over a number of assessments. Some of these packages have graphing facilities. Alternatively, some CAA and VLE systems will allow you to export the data to a spreadsheet or statistical analysis package.

Glossary

CAA	Computer-Assisted Assessment, ie using computers to assist in all processes involved in assessment
COLEG	Colleges Open Learning Group
Discrimination	A measure of the ability of an item to distinguish between candidates of different abilities
Distractor	An incorrect option in an objective item such as an MCQ
Facility	A measure of the difficulty of an item
Interoperability	The ability of software systems to exchange data in a commonly understood format
IMS	IMS Global Consortium Inc, the body responsible for the QTI specification (see below)
IP address	Internet Protocol address, ie the unique address which is sometimes given to computers which are connected to the internet
ISP	Internet Service Provider, allowing users to connect their machines to the internet
Item	The smallest separately identified question or task within an assessment
JISC	The Joint Information Systems Committee, an organisation which advises the UK further and higher education sectors on issues relating to the use of information and communication technologies
Key	The correct option in an objective item such as an MCQ
MCQ	Multiple Choice Question, the most commonly used item type in online assessment
Option	A possible response to an objective item such as an MCQ
QTI	Question and Test Interoperability, a specification which allows different software systems to exchange assessment data such as questions, assessments, responses and results
Reliability	How consistently an assessment is marked by different markers, with different candidates and over time
SCQF	Scottish Credit and Qualifications Framework
SFEFC	Scottish Further Education Funding Council
Stem	The question or statement to be responded to in an item
Validity	How well an assessment measures what you are trying to assess
VLE	Virtual Learning Environment, a software system which allows certain learning activities such as discussions, assessment and the presentation of materials to be carried out online
XML	A standard for the delivery of structured information over the internet

Further reading

Bloom, BS, et al, (1956) *Taxonomy of Educational Objectives: the classification of educational goals* volume 1 Longman, London.

British Standards Institution (2002) *BS7988: Code of practice for the use of information technology (IT) in the delivery of assessments*

Bull, J, Collins, C, Coughlin, E, Sharp, D (2001) *Technical Review of Plagiarism Detection Software Report*, JISC

Bull, J, McKenna, C (2001) *Blueprint for Computer-assisted Assessment*, University of Luton, www.caacentre.ac.uk

Burstein, J, Leacock, C, Swartz, R (2001) *Automated Evaluation of Essays and Short Answers*, Proceedings of the 5th International CAA Conference, Loughborough University, www.caaconference.com

Carroll, J, Appleton, J (2001) *Plagiarism: A Good Practice Guide*, JISC, www.jisc.ac.uk/pub01/brookes.pdf

Low, B, Sclater, N (2003 — in publication) *Authoring Reusable Computer-Assisted Assessment*, LTSN

Mitchell, T, Russell, T, Broomhead, P, Aldridge, N (2002) *Towards Robust Computerised Marking of Free Text Responses,* Proceedings of the 6th International CAA Conference, Loughborough University, www.caaconference.com

Race, P, *Designing Assessment to Improve Physical Sciences Learning*, LTSN Physical Sciences Centre

Ravet, S, Chevalier, P (2002) Assessment in an e-learning environment, *Prometeus Newsletter* No 21, www.prometeus.org

Sclater, N, Howie, K (2003) 'User requirements of the "ultimate" online assessment engine', *Computers and Education*, Vol 40, No 3, Elsevier Science Ltd, pp.285–386

SQA (2001) *Assessment and Quality Assurance for Open and Distance Learning* (A1030)

SQA (2003) *Guide to Assessment and Quality Assurance for Colleges of Further Education* (A0841/3)

SQA (1998) *Quality Assurance Principles, Elements and Criteria* (A0798)