AN INTERNATIONAL ASSOCIATION FOR MEDICAL EDUCATION

Online eAssessment

Reg Dennick
Simon Wilkinson
Nigel Purcell

AMEE GUIDE
Assessment
39

AMEE Guides in Medical Education

www.amee.org

Welcome to AMEE Guides Series 2

The AMEE Guides cover important topics in medical and healthcare professions education and provide information, practical advice and support. We hope that they will also stimulate your thinking and reflection on the topic. The Guides have been logically structured for ease of reading and contain useful take-home messages. Text boxes highlight key points and examples in practice. Each page in the guide provides a column for your own personal annotations, stimulated either by the text itself or the quotations. Sources of further information on the topic are provided in the reference list and bibliography. Guides are classified according to subject:

- Teaching and Learning
- Research in Medical Education
- Education Management
- Curriculum Planning
- Assessment
- Theories in Medical Education

The Guides are designed for use by individual teachers to inform their practice and can be used to support staff development programmes.

'Living Guides': An important feature of this new Guide series is the concept of supplements, which will provide a continuing source of information on the topic. Published supplements will be available for download.

If you would like to contribute a supplement based on your own experience, please contact the Guides Series Editor, Professor Trevor Gibbs (tjg.gibbs@gmail.com).

Supplements may comprise either a 'Viewpoint', where you communicate your views and comments on the Guide or the topic more generally, or a 'Practical Application', where you report on implementation of some aspect of the subject of the Guide in your own situation. Submissions for consideration for inclusion as a Guide supplement should be a maximum of 1,000 words.

Other Guides in the new series: A list of topics in this exciting new series is given below and continued on the back inside cover.

30 Peer Assisted Learning: a planning and implementation framework

Michael Ross, Helen Cameron (2007)
ISBN: 978-1-903934-38-8

Primarily designed to assist curriculum developers, course organisers and educational researchers develop and implement their own PAL initiatives.

31 Workplace-based Assessment as an Educational Tool

John Norcini, Vanessa Burch (2008)
ISBN: 978-1-903934-39-5

Several methods for assessing work-based activities are described, with preliminary evidence of their application, practicability, reliability and validity.

32 e-Learning in Medical Education

Rachel Ellaway, Ken Masters (2008)
ISBN: 978-1-903934-41-8

An increasingly important topic in medical education – a 'must read' introduction for the novice and a useful resource and update for the more experienced practitioner.

33 Faculty Development: Yesterday, Today and Tomorrow

Michelle McLean, Francois Cilliers, Jacqueline M van Wyk (2010)
ISBN: 978-1-903934-42-5

Useful frameworks for designing, implementing and evaluating faculty development programmes.

34 Teaching in the clinical environment

Subha Ramani, Sam Leinster (2008)
ISBN: 978-1-903934-43-2

An examination of the many challenges for teachers in the clinical environment, application of relevant educational theories to the clinical context and practical teaching tips for clinical teachers.

35 Continuing Medical Education

Nancy Davis, David Davis, Ralph Bloch (2010)
ISBN: 978-1-903934-44-9

Designed to provide a foundation for developing effective continuing medical education (CME) for practicing physicians.

36 Problem-Based Learning: where are we now?

David Taylor, Barbara Miflin (2010)
ISBN: 978-1-903934-45-6

A look at the various interpretations and practices that claim the label PBL, and a critique of these against the original concept and practice.

37 Setting and maintaining standards in multiple choice examinations

Raja C Bandaranayake (2010)
ISBN: 978-1-903934-51-7

An examination of the more commonly used methods of standard setting together with their advantages and disadvantages and illustrations of the procedures used in each, with the help of an example.

38 Learning in Interprofessional Teams

Marilyn Hammick, Lorna Olckers, Charles Campion-Smith (2010)
ISBN: 978-1-903934-52-4

Clarification of what is meant by Inter-professional learning and an exploration of the concept of teams and team working.

39 Online eAssessment

Reg Dennick, Simon Wilkinson, Nigel Purcell (2010)
ISBN: 978-1-903934-53-1

An outline of the advantages of on-line eAssessment and an examination of the intellectual, technical, learning and cost issues that arise from its use.

40 Creating effective poster presentations

George Hess, Kathryn Tosney, Leon Liegel (2009)
ISBN: 978-1-903934-48-7

Practical tips on preparing a poster – an important, but often badly executed communication tool.

41 The Place of Anatomy in Medical Education

Graham Louw, Norman Eizenberg, Stephen W Carmichael (2010)
ISBN: 978-1-903934-54-8

The teaching of anatomy in a traditional and in a problem-based curriculum from a practical and a theoretical perspective.

42 The use of simulated patients in medical education

Jennifer A Cleland, Keiko Abe, Jan-Joost Rethans (2010)
ISBN: 978-1-903934-55-5

A detailed overview on how to recruit, train and use Standardized Patients from a teaching and assessment perspective.

43 Scholarship, Publication and Career Advancement in Health Professions Education

William C McGaghie (2010)
ISBN: 978-1-903934-50-0

Advice for the teacher on the preparation and publication of manuscripts and twenty-one practical suggestions about how to advance a successful and satisfying career in the academic health professions.

44 The Use of Reflection in Medical Education

John Sandars (2010)
ISBN: 978-1-903934-56-2

A variety of educational approaches in undergraduate, postgraduate and continuing medical education that can be used for reflection, from text based reflective journals and critical incident reports to the creative use of digital media and storytelling.

Institution/Corresponding address:

Professor Reg Dennick, Assistant Director of Medical Education, University of Nottingham,
Queen's Medical Centre, Nottingham NG7 2UH, UK

Tel: +44 (0)115 9709788

Fax: +44 (0)115 9709922

Email: reg.dennick@nottingham.ac.uk

The author:

Reg Dennick is a biochemist turned medical educator and is Assistant Director of Medical Education in the Medical School at the University of Nottingham. He is also the Course Director for the University of Nottingham's Masters in Medical Education programme and is an educational consultant to the European League Against Rheumatism (EULAR), helping to design teaching and assessment materials for its on-line course for trainee rheumatologists. In 1999, in collaboration with Dr Simon Wilkinson, he led the design and development of the Network Learning Environment (NLE) project, which was the University of Nottingham's first purpose built virtual learning environment. He has overseen the growth and evolution of the NLE and been involved in developing the Touchstone assessment system which is now routinely used for high-stakes on-line summative medical examinations in Nottingham.

Simon Wilkinson has been involved with the use of IT in higher education since starting a PhD in hypertext and cognitive styles at Napier University, Edinburgh in 1995. In 1999 an opportunity presented itself to work on the TLTP3-86 project which was to build the University of Nottingham Medical School's first virtual learning environment (VLE). In 2002 he initiated a programme to build a dedicated assessment and survey system now called TouchStone which closely supports the pedagogic strategy of the Medical School. He has subsequently had considerable experience of running on-line summative assessments and is skilled in managing this process. Today he continues to oversee strategic developments of both the NLE and TouchStone focussing on issues such as standard setting and curriculum mapping.

Nigel Purcell currently works as the Senior Education Advisor at the Higher Education Academy Subject Centre for Medicine, Dentistry and Veterinary Medicine, based at the University of Newcastle. In that capacity he manages the workshop programme and has specific responsibility for supporting the development of medical, dental and veterinary educators. Prior to that he was Staff Development officer in the Faculty of Medicine at the University of Liverpool where he was responsible for delivering a wide range of teacher training courses in support of the PBL based curriculum delivered there. Before entering the field of medical education, Nigel worked for a number of years as an educational consultant in the field of vocational education in further education. He has considerable experience of teaching in both further and higher education and his main interests are in how to support staff involved in educating health care educators and in the specific characteristics of teaching and learning in the clinical environment.

Guide Series Editor: Trevor Gibbs (tjg.gibbs@gmail.com)
Published by: Association for Medical Education in Europe (AMEE), Dundee, UK
Designed by: Lynn Thomson

ISBN: 978-1-903934-73-9

Contents

Abstract

In this Guide the authors outline the advantages of on-line eAssessment and examine the intellectual, technical, legal and cost issues that arise from its use. The Guide outlines the major assessment types that are suitable for on-line assessment and makes a key distinction between formative and summative assessment. The focus is primarily on the latter since that is where the difficulties are most acute and robust systems most critical.

A range of practical issues relating to the key stages in running a summative e-Exam are explored and advice given on system requirements and on how to ensure that the exam runs smoothly when you "*go live*". This section includes consideration of the way that using eAssessment might affect the standard setting and results analysis process

The section on future trends in on-line assessment explores such possibilities as computer adaptive testing and the automated assessment of free text answers. Finally there is a consideration of the implications of these trends for management.

TAKE HOME MESSAGES

- eAssessment offers substantial potential benefits but needs to be carefully managed to minimise risks
- Fundamental assessment principles such as ensuring that assessment instruments are reliable and valid are just as important in eAssessment
- eAssessment used formatively offers rapid and effective feedback to learners and can be used to substantially enhance the learning process
- The risks of eAssessment are greatest in the context of summative assessment, so make sure you have adequate hardware and back-up systems when running summative exams
- eAssessment offers the potential for new types of question and formats which can be used to enhance reliability, validity and utility

Introduction

The use of computers and information technology (C&IT) is now well established in medical education and forms the subject of electronic learning or eLearning (McKendree, 2006). Learning is conceptually linked to assessment, where the amount and quality of learning is measured for reasons of safety, grading or feedback. Thus one aspect of eLearning is electronic assessment or eAssessment and the purpose of this Guide is to outline the main features of eAssessment and the methods that are being used to implement it. Assessment is traditionally divided into formative assessment and summative assessment and the eAssessment variants of these will be described. However, the bulk of this guide will be devoted to summative eAssessment as that is where the greatest practical challenges lie and where some of the primary advantages of this technology can be found.

Historically eAssessment was always associated with the development of eLearning. Some of the earliest forms of computer assisted learning (CAL) were frequently just 'drill and practice' programs using multiple choice questions, sometimes with feedback or branching algorithms that could respond to individual choices. eAssessments have therefore largely developed from conventional forms of 'objective' assessment so that paper based versions of multiple choice, true-false-abstain, multiple, response and extended matching questions have been converted into electronic versions. However, once this process has occurred, a number of opportunities and advantages become apparent which can transform assessment and make it a much more relevant, valid, exciting and meaningful process.

Historically eAssessment was always associated with the development of eLearning.

Some of these opportunities and advantages will be discussed below as well as some disadvantages and practical difficulties that derive from the computer based medium itself. This Guide will concentrate on computer-based assessment using a client-server architecture such as the Internet and the use of computer-based assessment for objectively marked items. The use of computers to assess or evaluate significant amounts of text will not be covered, however, and the reader is referred to Valentini et al 2003 for more information on this type of assessment.

For the purposes of this Guide we will assume that readers are familiar with the creation of high quality, reliable and valid assessment items. Readers are referred to the following resources for references: Case & Swanson (2002) and Holsgrove & Elzubeir (1998).

Advantages of on-line assessment

Before looking at formative and summative eAssessment in more detail it is worth outlining some of the general arguments for the use of online assessments (Oblinger, 2006; Sim et al, 2004) and some of the key principles of assessment that apply to these situations.

Students entering higher education today, typically:

- have experience of computer technology in both their school and home lives
- expect interaction
- want a visual experience
- desire rapid feedback on their activities
- want technologically modern courses
- want a more holistically challenging assessment environment.

From the point of view of teaching and administration staff, the move to assessing students online also offers a number of advantages:

- Online assessment can reduce marking loads
- Results can be available as soon as an exam is finished
- Results can be immediately reviewed by an exam board
- A variety of online quality checks can be performed

Box 1 is an outline of how on-line assessments can potentially fulfil these demands plus the disadvantages that need to be considered.

From the point of view of teaching and administration staff, the move to assessing students online also offers a number of advantages.

Box 1

Computer-Based Assessment – Pros	Computer-Based Assessment – Cons
Students: • Easily monitor their academic progress by means of formative papers with feedback, available 24/7 • Answers can be entered/altered quickly and clearly • Assessments can be modified to accommodate special needs. • Interactive, adaptive and multimedia question types possible with high validity. Academic Staff: • Questions can provide a more valid and holistic way of assessing knowledge • Can monitor the assessments of students to personalise feedback. • External examiner can have instant access to a paper electronically. • Interactive, adaptive and multimedia question types possible. Administrative Staff: • Fast Marking – scales well with additional examinees • Saves paper	Costs: • Assessment system software licence • Powerful servers • Large numbers of clients (PCs) • Staffing • Physical environment – large air-conditioned labs. User Training: • Students must learn how to use the assessment system (should be during formative assessments) • Staff must be trained in how to enter questions and utilise the full capabilities of the system Risks: • External hackers/viruses • Internal staff/student security policies. • Failure – power / hardware / software

Issues in eAssessment

The importance of good assessment is highlighted in Boud's (1995) statement, "*Students can, with difficulty, escape from the effects of poor teaching, they cannot… escape the effects of poor assessment*". This principle applies equally to eAssessment and the work of the UK Collaboration for a Digital Repository (UKCDR) (2007) and Schuwirth & van der Vleuten (2006) can be combined to create four broad perspectives with which it should be possible to defend any form of assessment in the following areas:

- Intellectual
- Legal
- Technological
- Economic

"Students can, with difficulty, escape from the effects of poor teaching, they cannot… escape the effects of poor assessment"

Intellectual Issues

Summative assessments can be used for high stakes decision-making processes. Given such importance it is critical that the effect of utilising eAssessment on the *reliability* and *validity* of the assessments is considered.

Summative assessments can be used for high stakes decision-making processes. Given such importance it is critical that the effect of utilising eAssessment on the reliability and validity of the assessments is considered.

- **Reliability**: The reliability of an assessment refers to its ability to consistently give the same measure of learning when used repeatedly despite sampling error. The most common cause of un-reliability in testing is a lack of consistency in the use of assessment criteria by a marker. In the sort of objective testing we are describing here, where objective criteria are decided beforehand and questions are marked electronically, this type of reliability problem is diminished.

  However, another form of reliability is the internal consistency of the assessment task, usually measured by correlating individual item scores to other items or to the global test score which can be processed to give a value of reliability such as Cronbach's alpha (Cronbach, 1951). Because with on-line assessments it is possible to supply a different set of questions from a question bank to different individuals in the same examination, or to generate different numerical values for calculations or problem solving items within a question the questions delivered to individuals can vary slightly. Provided the range of these variables is within agreed boundaries overall the reliability of the test should not be greatly compromised.

  Reliability can also be influenced by learners' personal factors such as their propensity to guess, whether they have dyslexia or how easily they are fatigued by using a Visual Display Unit (VDU). The influence of these factors on reliability will be discussed later.

- **Validity**: In general, assessment validity is concerned with whether an assessment measures what it is designed to measure and can be sub-divided into a variety of different types (Dent & Harden, 2005):
  - *Content validity*: 'does the test measure and sample relevant learning objectives or outcomes?'
  - *Construct validity*: 'does the test measure an underlying cognitive trait eg. intelligence?'
  - *Concurrent validity*: 'does the test correlate with the results of an established test?'

– *Predictive validity*: 'does the test predict future performance?'
– *Face validity*: 'does it seem like a fair test to the candidates'

The most important elements that might be influenced by being online would be content validity and possibly the related concept of construct validity. However, Schuwirth & van der Vleuten (2006) argue that assessments must also have face validity for students. This is an important issue particularly when introducing on-line eAssessment for the first time to students who may be unfamiliar with its processes and may require reassurance.

The most important elements that might be influenced by being online would be content validity and possibly the related concept of construct validity.

Certainly content validity can be enhanced and expanded by means of online assessment technology. For example the following additional features can be added to online questions:

– animations, video and sound (if headphones are used in the examination room).
– 'Hotspot' questions which require students to place a mark anywhere on an image or diagram
– dragging labels directly over an image.
– simulations

A system that can test the students in a realistic scenario through a number of stages is the Tripartite Interactive Assessment Delivery System (TRIADS) created in a partnership between the University of Liverpool, University of Derby and the Open University in the UK (TRIADS, 2007). Assessments are created in Authorware and are tailor-made for each question.

In all these cases the online nature and technological aspects of the assessment can significantly influence the authenticity of questions that can be created in comparison to other forms of paper-based assessment media (Sim et al., 2005). Evidence for increased validity can be found in an evaluation of multimedia online examinations by Liu et al. (2001). They investigated student and staffs' attitude to multimedia exams and found very strong support for their use. For example they found that:

- assessment more closely matched the material that was being taught
- the presentation of more than one medium of information seemed to aid the students' recall
- questions reflected real-world situations more accurately
- students seemed to learn more in these assessments, which helped them as they continued their studies.

Legal Issues

The legal issues for an online examination system are:

- copyright for graphics, video or sound
- questions from other institutions

If an online exam uses graphics, video or sound the copyright for these materials must be obtained for them to be used in the system, especially if they are to be archived on the system for some time after the exam or possibly reused in further exams. Related to this, there is also the possibility that academic staff may bring questions with them from other institutions that may still belong to those institutions rather than the individual and conversely, take material away with them if they leave. A 'Take Down' policy needs to be in place in case materials with such issues are discovered in use.

Technical Issues

This Guide concentrates on discussing the issues surrounding one of the most popular types of assessment architectures: client-server. This is the classic Internet architecture whereby an end user sits at a personal computer, the client, and requests pages to be sent from a website, the server. However, what constitutes the assessment 'system' is more than the assessment software. It includes additional sub-systems such as routers, switches, and network load-balancers, plus a range of operating system and applications software. It is important to understand how these various sub-systems relate to one another and what would happen to the examinees if one or more sub-systems failed. The ability to recover from a technical failure is one of the key issues of conducting online exams and disaster recovery must be planned in advance. Although the literature surrounding high profile summative failures is rather sparse, Harwood (2005) presents a frank account of the processes the University of Southampton followed after one of their assessment systems failed catastrophically.

The ability to recover from a technical failure is one of the key issues of conducting online exams and disaster recovery must be planned in advance.

Technical and practical issues will be further discussed in the later section on 'Exam Delivery'.

Economic Issues

It is a common fallacy to assume that online assessment will be cheaper than alternative forms simply because a whole cohort can be marked in a matter of seconds. However the following costs need to be taken into consideration:

It is a common fallacy to assume that online assessment will be cheaper than alternative forms simply because a whole cohort can be marked in a matter of seconds.

- large numbers of computers are required for a simultaneous start
- additional invigilators will be required if these machines are located in different computer rooms
- dedicated assessment servers are required to minimise failure risk
- assessment software
- departmental/institutional staff required to support the system
- educationalists advising on pedagogic approach and assessment strategies
- programmers' salaries
- trainers familiar with the assessment software
- IT support technicians

Some of the costs of online assessment are considerable: thousands of pounds can be spent on server hardware, potentially large computer labs, plus the license cost of the assessment software itself. These costs may be

off-set by the fact that computer labs have other uses and assessment software often includes survey and questionnaire software. Then there are less tangible aspects to costs such as members of IT support staff spending more time maintaining systems. On the other hand, compared with Optical Mark Reader (OMR)-based assessment, online systems can mark substantially faster, more accurately and can save paper and printing costs. A complete and comprehensive auditing of all these costs would be useful in the justification of online assessment. Of course, final decisions regarding whether to use online or offline assessment will include additional factors such as the quality of the assessments that can be created.

Having dealt with some key eAssessment issues and concepts it is now time to look briefly at some core assessment principles in the context of eAssessment and then look at how it can be used in the context of formative and summative assessment.

Assessment Types

MCQs and EMIs

It is assumed that readers are familiar with the major objective formats of Multiple Choice and Extended Matching as outlined in the Guide produced by Case & Swanson (2002). These formats are employed in most conventional types of assessment and are readily modified for the on-line environment by including images and even video clips.

Fill in the gap (Cloze) and text/number entry

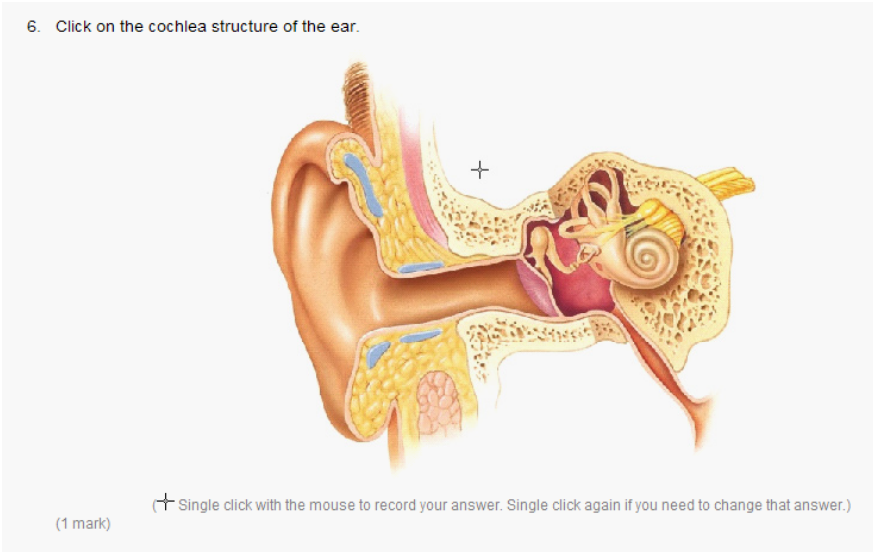
These are related systems that involve the student entering single words, phrases or numbers into a section of text or a designated text/numerical box. Cloze is the technical term for inserting deleted words into a section of text in order to complete it correctly and hence for assessing recall of factual information (Taylor, 1953). Single words, phrases or numbers can be inserted into designated boxes as answers to a variety of question types. The effectiveness of solutions to the problems of error trapping the input and recognising correct answers from all possible inputs is a limiting factor in the use of this question format.

Image hotspots [Figure 1]

Image hotspot type questions are good at assessing visual knowledge that would be difficult to achieve though an MCQ or other textual question type. They have a second advantage in that there are no visual cues as to where the correct answer lies, there are no discrete distractors to choose from, and each pixel is a potentially correct or incorrect answer.

Image hotspot type questions are good at assessing visual knowledge that would be difficult to achieve though an MCQ or other textual question type.

FIGURE 1
Image hotspots

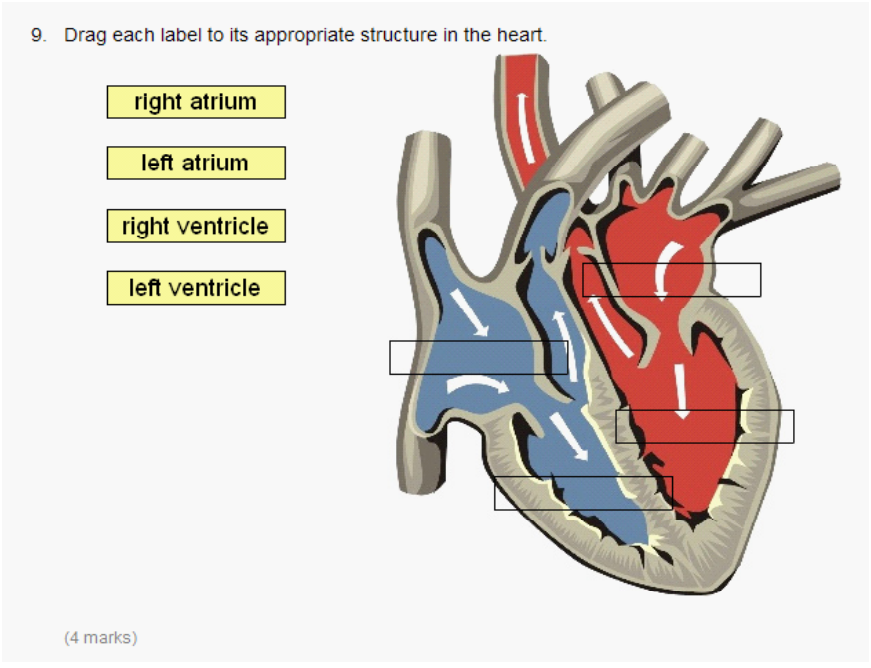


Labelling (drag 'n drop) [Figure 2]

Labelling questions, like image hotspots, are ideally suited to assessing visual knowledge, they differ in the cues they provide. With a labelling question a number of 'place holders', the empty rectangles, are pre-displayed over the image of interest. The examinee must drag labels from the left and drop them into the relevant place holders. Sometimes a larger number of labels than placeholders are used to make the question more difficult.

Labelling questions, like image hotspots, are ideally suited to assessing visual knowledge, they differ in the cues they provide.

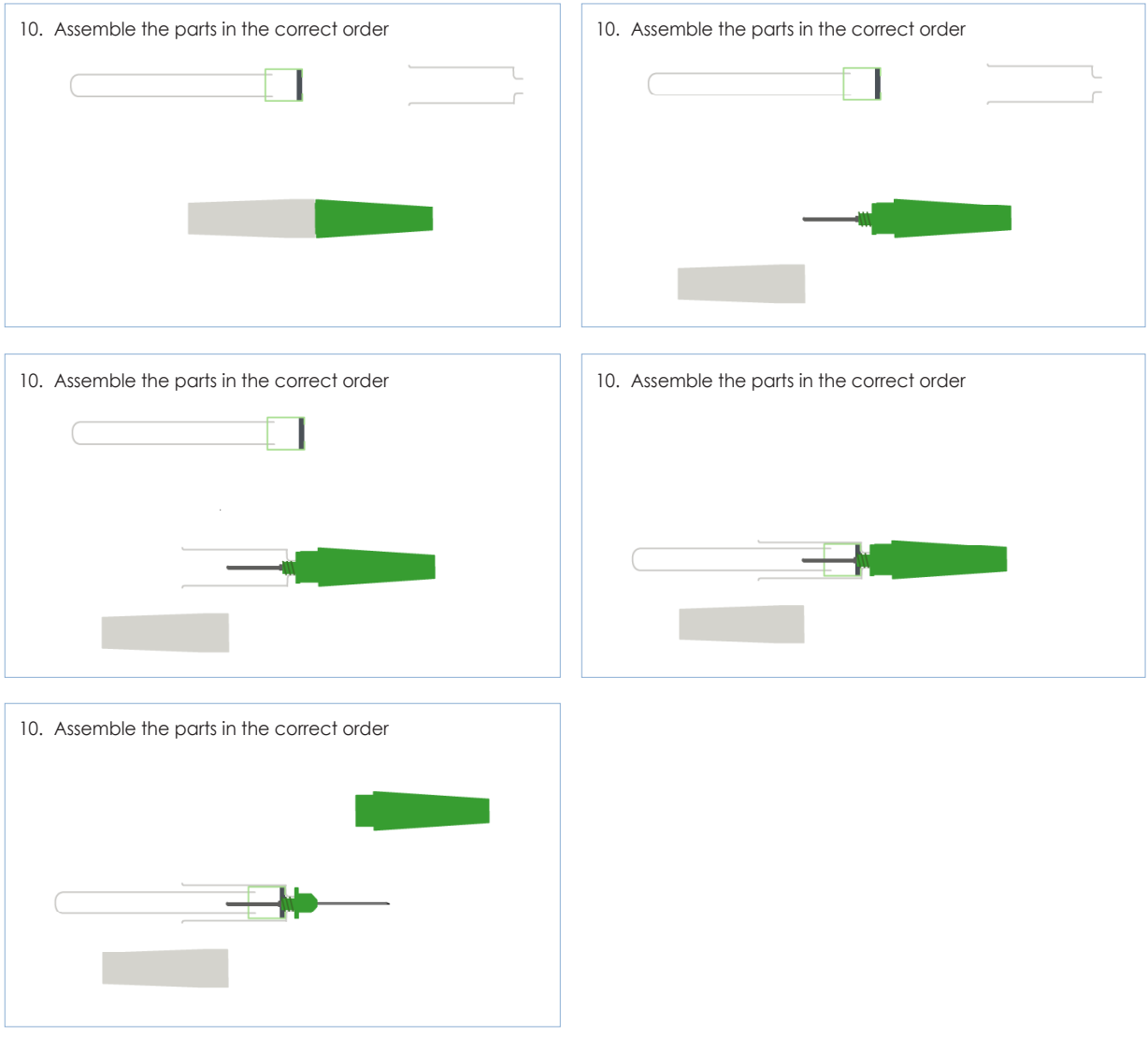
FIGURE 2
Labelling (drag and drop)



Simulations [Figure 3]

Certain systems such as Perception and TouchStone can accept questions built externally using languages such as Flash. The screenshot below shows one such Flash example which simulates the process of setting up a Vacutainer for taking a blood sample. This is testing the examinee's procedural knowledge through direct mouse interaction with a simulated Vacutainer. The student has to complete the virtual process by clicking and dragging items in the correct sequence. The simulation is programmed with a physics model whereby each separate element can hit or be attached to other elements rather than passing over or under the other object.

FIGURE 3
Simulations



Other simulations that can in principle be used as assessments include SimMan (SIMMAN, 2007) and variants on this technology. Although currently this is not a client-server technology the principle of interacting with a human physiological simulation can be used as a form of assessment.

Video

The ability to deliver video or moving images to a student during an assessment considerably extends the scope of question formats. Videos of patients, doctor-patient interactions, procedures, consultations and communications can all be used to create appropriate assessment scenarios that have high content validity. Video can be used to set up a scenario which can be subsequently assessed by means of the formats described above.

Formative eAssessment

Formative assessment involves assessing students directly in the context of learning in order to give them feedback on their progress. It may involve direct observation of student behaviour and the giving of oral feedback or it may involve giving students problems, assignments or even exam questions to take under informal conditions followed by feedback on performance. The aim is to allow students to monitor their progress as they are learning in order to improve their learning (Wood, 2007).

Formative assessment involves assessing students directly in the context of learning in order to give them feedback on their progress.

The on-line environment is ideally suited to this form of assessment as it is relatively straightforward to provide students with access to a variety of self-assessments including online past papers that can be taken in their own time under non-examination conditions and which can give them feedback on their progress. The variety of assessment types available are identical to what can be used for on-line summative assessments.

On-line formative eAssessments can be provided at the end of teaching sessions or episodes to consolidate student learning. They may be embedded in Reusable Learning Objects (RLOs). On the other hand formative eAssessments might take the form of past examination papers that students can take during the academic year to test their progress and familiarise themselves with the types of questions they might receive in summative exams at the end of a module or year. A useful policy is to make a formative version of all summative eAssessments available to students so that they can familiarise themselves with the formats and levels of questions. Building feedback into questions makes them a useful learning resource.

An interesting example of formative eAssessment is on-line Progress Testing where students undertake a series of on-line assessments during the year that sample questions from the whole of the course. Students are given a period of time such as a week to take the test after which they are given a mark and the average mark for the year. They are then allowed to go back into the on-line assessment to see which questions they answered correctly or incorrectly and to read the feedback comments built into the questions. Although it can be compulsory to take the test, giving it a quasi-summative flavour, the mark they receive is not used in any formal sense and only serves to give them some feedback on how well they are progressing through the course.

Students can take formative assessments in their own time without elaborate security and without the need for invigilation. The problems encountered when online assessments become summative and hence high stakes will be discussed in the next section.

Summative eAssessment

In a review looking at medical education Cantillon et al (2004) found the use of computers for summative assessment much more limited. Factors preventing wider adoption of online summative assessment included lack of space and security concerns. The publication of failures (Smailes, 2002; Harwood, 2005; Heintz & Jemison, 2005) also does little to reassure the unconverted. A key aim of this Guide is to provide information to those wishing to implement the use of online summative assessments and it is this aspect of eAssessment on which we now wish to focus.

This section will look at the examination cycle, the planning, creating and implementing sequence that needs to be undertaken to create successful summative eAssessment. It will then look in detail at the issues surrounding the delivery of the examination into a summative environment as this is where novel problems can arise that users need to be aware of and to have plans to deal with.

The examination cycle

Room Bookings

Such rooms should ideally be large enough to examine the entire cohort simultaneously or through two sittings. For many Universities and Medical Schools this can be a major problem. Summative eAssessment is a recent phenomenon and the infrastructure required is not necessary available for the large cohort sizes that exist. Booking in good time is important due to pressure from other departments to reserve the same spaces. Once a booking is confirmed students should be notified of the computer lab details, often through a posting on a virtual learning environment or portal. In situations where a cohort has to be split into two to be examined certain additional steps must be covered. For example, a list of which students have been assigned to each group is necessary.

Item Development

The advantage of using server-based assessment systems is that it is very easy to collaborate when developing items without physically meeting the other question setters. When working in complex domains it is likely that multiple authors will wish to author items for a single exam. In such cases the assessment software should support some sort of group or team working and be able to stop editing conflicts.

The advantage of using server-based assessment systems is that it is very easy to collaborate when developing items without physically meeting the other question setters.

However, when using the 'stateless' architecture of the web it is very easy for one author to inadvertently overwrite the changes made by a different author who is working unbeknown to the first author at the same time. Some systems can prevent this situation from occurring by effectively placing a 'lock' when the first author goes into an item for editing. Any subsequent authors are informed that the item is locked and that they will only be presented with a read only version. Automatic audit trails are also useful so that in the event of problems with a question it is easy to look back through a change log.

Item storage

Establishing a 'deletes' policy is good practice when dealing with mature question banks. Some assessment systems will produce errors if a member of staff wishes to run a report on a student cohort who took an exam some time ago that uses one or more questions which have been deleted from the bank. Many disciplines are periodically inspected by governing professional bodies and increasingly institutions are providing guest accounts for these institutions to log into virtual learning environments (VLEs) and online assessment systems. In the past it has been relatively easy to find past data filed carefully by year within physical filing cabinets, but moving all this information into the electronic domain raises additional concerns associated with the security of electronic data.

A reliable and regular backup of an assessment system (questions items, papers, user accounts and past exam results) should be made, ideally to a separate and secure location away from the primary assessment server. An archive of backups is also invaluable if past data that has been deleted also needs to be retrieved. Just as the quality assurance process should be periodically tested, so too should the backup procedures.

A reliable and regular backup of an assessment system (questions items, papers, user accounts and past exam results) should be made, ideally to a separate and secure location away from the primary assessment server.

In addition to appropriate hardware, the capabilities of the assessment software can play a key role in item storage. Each system is capable of storing pieces of data such as the question lead-in and options that form part of the question, but it is also important to be able to store associated meta-data. This meta-data will not be seen by the students during exam delivery but makes overall staff administration of large question banks easier. The amount of metadata stored will differ between assessment systems but most will include the following types for each question:

- Name of question author
- Time/date item was created
- Time/date item was last edited
- Keywords
- Difficulty level (e.g. Bloom's Taxonomy)

Item selection

Excluding adaptive assessment systems, there are two distinct methods of creating papers from items in a question bank. The simplest option, as with a paper-based exam, is for the exam authors to specifically select which questions will be used and the order in which they will be listed. A more complex method utilises the power of the computer to randomly pick out questions from the bank. Two sub-types of randomisation are possible:

1) all examinees receive the same questions within the exam but the order of presentation is randomised, and
2) the questions used on an exam paper are randomised such that different examinees will answer slightly different question sets.

This latter type of randomisation is often favoured for reducing plagiarism as neighbouring students will have different questions. However, their use in summative examinations raises issues of exam paper comparability and

hence reliability. However, as previously mentioned if all questions are of equal standard and are aimed at the same constructs this should not be a major problem.

Item testing

For the purposes of this Guide we will assume that general quality assurance mechanism exist which can deal with the creation and use of assessment items. However, online assessment systems create additional problems that need to be dealt with.

Where possible all quality reviews should be done online using the same assessment software as will be used to deliver the final assessment to students. The most common problem to slip through review processes that the current authors have witnessed is formatting issues that have arisen when, for example, a member of staff copies and pastes an original question from a word-processor into the target assessment system.

Before running a summative examination online it is useful to perform the following tests that will detect problems in the marking routines:

- Do not answer any items: score should be zero
- Answer all items correctly: percentage score should be 100%
- Answer all items incorrectly: score should be zero.

Setting the pass mark and standard setting

The pass mark for an examination can be set in a number of ways. Norm-referencing involves setting a pass mark after the examination has been taken which allows a previously decided proportion of students to pass the exam. In general this method is no longer recommended for a variety of reasons, not the least of which is its intrinsic unfairness; students pass or fail not on their own merit but depending on how the overall cohort does. However, it can be used in high stakes examinations when there is a restriction on the numbers of students who are able to pass on to the next phase or as part of an entrance exam with a limited number of places.

The commonest and fairest method of setting a pass mark is criterion-referencing which involves setting a fixed pass mark initially and allowing any students who exceed it to pass. Many Universities have regulations that prescribe fixed pass marks, such as 40%. Historically UK and US Medical Schools employed large numbers of true/false questions, with negative marking to inhibit guessing and the option of abstaining. However, these types of questions are no longer recommended (Case & Swanson 2002) and have all but died out to be replaced with a broader spectrum of question types: extended matching, single best answer, multiple response, ranking and image hotspots as described earlier. For examinations constructed of these types of questions students are instructed to answer all questions and not to abstain hence there is a possibility that a correct answer will be selected merely by chance.

The commonest and fairest method of setting a pass mark is criterion-referencing which involves setting a fixed pass mark initially and allowing any students who exceed it to pass.

One way of dealing with this random factor suggested by Harper (2003) is to incorporate a 'correction for guessing' at the post-exam grading stage. This is the total mark that would be obtained by chance alone which can be calculated from the summed probabilities for each type of objective question within the test. For example each component of a five stem multiple choice (MCQ) question marked out of one would have an expected mark equal to its probability of 0.2. This correction for guessing can then be subtracted from the total mark and used to rebase the assessment and calculate a corrected for guessing score. The process is analogous to subtracting the 'noise' from a set of data in order to more easily see the 'signal'. Harper describes using a spreadsheet for this purpose. However, some assessment systems (for example TouchStone, 2007) can perform such calculations automatically. Of course a pass mark still has to be applied to the rebased exam data which may be fixed by University regulations or the alternative approach of standard setting adopted.

The process of standard setting has recently been reviewed by Norcini (2003). Essentially the method uses teams of subject-matter experts to discuss each item on a paper separately and to make some form of collective decision regarding how many 'borderline candidates' will answer the item correctly. There are a number of different techniques for doing this, Ebel (1972) and Angoff (1971) being two of the more common. Although both techniques do not explicitly take into account the probability of selecting a correct answer by chance, the overall calculated pass marks are usually significantly above what could be achieved through guessing alone hence the probability can be dismissed. Where possible it is recommended that an assessment system with built in support for standard setting is used when setting pass marks in this way. It is time consuming to set up spreadsheets to perform standard setting manually and there is always the risk that the questions may be inadvertently changed when copying from the assessment system into the spreadsheet or vice versa.

Exam Delivery – System Requirements

The specification of the client-side computers that the students will use during the exam is not problematic today; modern desktop computers have a surplus of power for running web-based exams. However, the server that hosts and serves each assessment is a different issue. Some basic features can be suggested for a successful fault-tolerant server hardware platform:

Reliability of the computer systems

When an online exam begins all the client computers that the students are using will send their requests back to a single web server which holds the exam paper. The main drawback of this client-server architecture is that it introduces a single point of failure. In practice there are a number of different things which can be done to minimise this risk.

When an online exam begins all the client computers that the students are using will send their requests back to a single web server which holds the exam paper. The main drawback of this client-server architecture is that it introduces a single point of failure.

With primary storage (RAM) error correcting code (ECC) modules can be specified on some servers to minimise errors that could crash software. In terms of secondary memory (hard disks), RAID 5 is a useful configuration. A RAID 5 arrangement requires a minimum of 3 separate hard disks to be

installed within the server and the reading and writing of data is spread across these disks with additional parity data being written in order to check for any errors in this process.

High-end servers will normally come supplied with two power supplies and two or more network connection ports. Where possible the two network connections should go to different switches on different parts of the network so that Internet traffic to and from the server can be routed even if one switch fails.

Finally, a large uninterruptible power supply (UPS) system should be installed which can power the server until either a backup generator starts or mains power is restored.

Storage

A server must have enough primary and secondary memory to support the maximum class size expected for an online assessment. The higher the number of simultaneous users the more primary memory (RAM) will be required to run the assessment.

Factors influencing secondary memory (hard disk) size include:

- amount of data that needs to be stored:
- amount of multimedia data used in questions,
- number of students at each exam
- total number of assessments planned for any given time period.

Performance

Although there are software applications which can be used to simulate exam load these should not replace real-world test sessions in non-critical (i.e. non-summative) periods. Heintz & Jemison (2005) stress the importance of benchmarking and simulating exam delivery. A good way of doing this is to hold one or more invigilated and compulsory formative exams with the same cohort that is scheduled to take the final summative exam. On the basis of these load-tests a couple of different strategies can be employed:

1) a staggered start of the examinees in blocks (Heintz & Jemison, 2005), or
2) starting the whole cohort simultaneously in a similar way to a paper exam if the system can respond fast enough.

Although there are software applications which can be used to simulate exam load these should not replace real-world test sessions in non-critical (i.e. non-summative) periods.

Independence

Where financially possible a dedicated assessment server should be used which is independent of other systems.

Where financially possible a dedicated assessment server should be used which is independent of other systems.

Exam Delivery – Going Live

The live delivery of an online summative exam, under conventional exam conditions is the most crucial phase of the process. If a system does not respond as expected a contingency plan must be put into place. Although disaster recovery will be covered later there can be no substitute for rigorous and comprehensive planning of the exam delivery stage. Three main issues dominate:

- security,
- software usability
- administration.

There is an international standard produced by the British Standards Institute entitled 'Code of practice for the use of information technology (IT) in the delivery of assessments' (BS ISO/IEC 23988, 2007) which covers many aspects of exam delivery in generic terms.

Security

The avenues for potential security breaches can be broken down into two broad categories: external security and internal security.

- **External Security**: Security risks are possible with any server attached to the Internet. Hackers anywhere worldwide are constantly using methods and software systems to root out vulnerable servers. When breached a hacker might crash the server and thereby stop an exam or use the assessment server to send out spam email which will affect its performance. Networking and security experts from the parent institution should be involved in the assessment process to ensure external loopholes are discovered and patched before the hackers can exploit them. This process is not simply an initial system setup activity but an ongoing virtual battle in cyberspace.

  A firewall (either hardware or software) is a system which controls requests and protocols accepted and transmitted by a server. Most assessment systems will require HTTP or ideally HTTPS (encrypted) protocols so a firewall can be used to deny access to other protocols such as FTP and email.

  All software sub-systems should be patched and kept up-to-date; this includes operating system (Windows, Linux, etc), web server (Apache, IIS, etc), and applications software which would include scripting languages (PHP, .NET, etc) and often a database (MySQL, Oracle, MS SQL, etc).

- **Internal security**: Usually a web server will deliver pages twenty-four hours a day to any computer worldwide but good assessment systems are able to limit access using any combination of course, module, year of study, time/date and room. A system should only deliver an online assessment to a relevant cohort of students studying a specific module, at the prescribed time and only to the examination room used.

  If two sittings of an exam are required through lack of computers students in the second group should not be able to log into the exam paper while the first group are taking the assessment. Students should not be able to leave early and inform students not yet examined what the questions are.

A system should only deliver an online assessment to a relevant cohort of students studying a specific module, at the prescribed time and only to the examination room used.

Two solutions are possible here.

1) The two groups are examined back-to-back with no one allowed to leave the examination room for either sitting,
2) Different examination papers are used for each group – either two manually created papers or the use of papers which randomly select questions.

The accommodation of individuals needing extra time should also be planned. Ideally, candidates with additional time, such as dyslexic students, should be examined in a separate computer lab. Where this is not possible then the complete additional period of time permitted for these students should only start after all students have left the room.

Who should have access?

Which staff have access to system wide privileges, who can add and alter questions and who can only run reports is another key security issue. Some assessment systems will utilise the authentication systems within an overall VLE architecture. Other systems will employ authentication such as Lightweight Directory Access Protocol (LDAP) to ensure that only registered users can access the assessment system. More proprietary or home-grown systems may even use their own maintained lists of authorised users. In the last case it is vital that key personnel are identified who are responsible for maintaining these lists every year as new students are registered with the institution. Whatever method of authentication is used two important conceptual issues have to be considered and decisions made:

1) identification – which individuals can access a system, and
2) authorisation – which parts of the system these individuals are allowed to access. For example, in terms of identification it could be all students and teaching staff connected with a particular course or module; however, in terms of authorisation the students will only be allowed to view and answer certain assessments at controlled times whereas staff will be able to add questions, edit, delete and run reports.

Preventing cheating

Even within a group of legitimate examinees who are allowed to access an online exam, security is still very important. The importance of summative examination leads some students to cheat. In a study of school and further education examination Underwood (2006, p.1) states, "*Although there remains some debate on whether the incidence of academic malpractice is increasing, it is widely acknowledged that it is a very significant problem*". Referencing the work of Hinman (2000) she suggests a three pronged approach to reducing academic malpractice, summed up as the three 'Es':

- Ethics
- Engineering
- Enforcement

Even within a group of legitimate examinees who are allowed to access an online exam, security is still very important. The importance of summative examination leads some students to cheat.

Ethics (the virtues approach)

This approach is based on the establishment of an agreed code of practice which can be circulated in a transparent process to both students and staff. Engineering (the prevention approach).

There are several steps which can be taken using the 'engineering' approach:

- Reduce recycling of past exam papers;
- Introduce seating plans, students sitting next to 'strangers' are less likely to cheat;
- Introduce visual barriers (see figure 4) where adjacent workstations are close (BS ISO/IEC 23988, 2007);
- Screen covers/modifiers which only allow the user to see the screen from a narrow range of angles perpendicular to the screen preventing adjacent students from observing another's screen.
- Limit the materials students may bring into the examination room;
- Secure browser (Heintz & Jemison, 2005) or desktop whereby students cannot use any other part of the computer's functionality other than the examination itself.
- Normal facilities such as email, access to the wider Internet and chat must all be disabled for the duration of the exam.

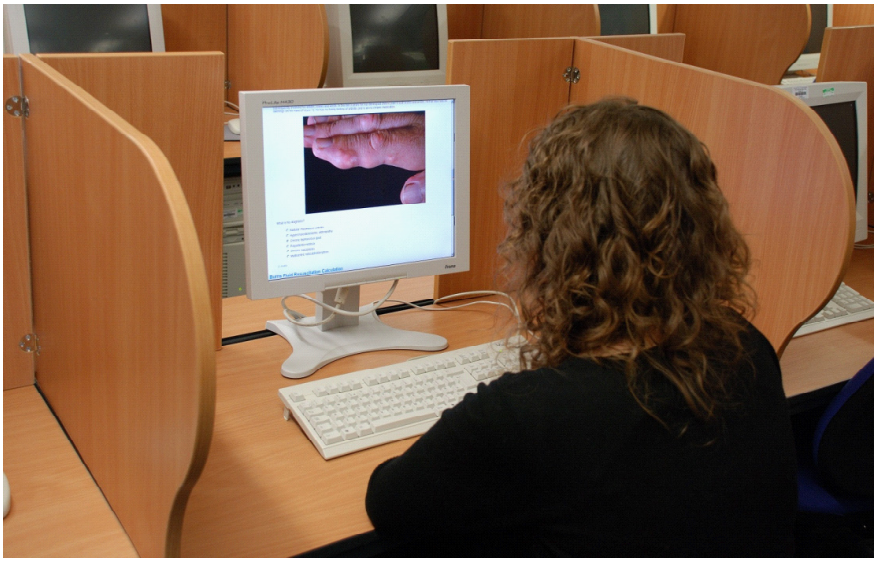
Enforcement (the police approach)

One enforcement approach is to use statistical analysis after an exam to detect when the answer patterns of two or more candidates are unlikely to be that similar by chance. Such techniques are then used with IP address recording and seating plans to see if the suspected individuals were physically in close proximity.

One enforcement approach is to use statistical analysis after an exam to detect when the answer patterns of two or more candidates are unlikely to be that similar by chance.

FIGURE 4

An example of physical barriers used to prevent cheating in a multi-purpose computer lab where adjacent workstations are close. These barriers may be taken down and stored when the lab is not required for assessments.



Software Usability

Usability is a second important aspect that should be one of the key factors used when deciding which assessment system to install. Students should receive a mark which reflects their level of subject matter understanding rather than their IT capabilities. The assessment system employed must effectively become transparent to the students. Nielsen (2005) lists 10 criteria which can be applied to any interactive software system to measure usability in a more objective manner. In addition to using systems with high usability it is important to ensure examinees are exposed to the software before any summative examinations so they have time to familiarise themselves. Formative assessments should be written in the same software as the final summative exam and made available to the students prior to the exam.

Students should receive a mark which reflects their level of subject matter understanding rather than their IT capabilities.

Special needs

It is necessary to identify any examinees with special needs. Most countries will have a form of legislation designed to protect the interests of users with special needs or disabilities. In the UK there is the Special Educational Needs and Disability Act (SENDA, 2001) which is now enshrined in law. Many institutions use 'accessibility units', or other place with similar titles, to provide centres for advice for students with particular requirements. Having clearly documented protocols and networks of support established is important so that these units can feed back to, in many instances, a school or faculty-based administrative unit that may then need to speak to an IT expert to establish what is and what is not possible to change for a student. Broadly speaking there may be four factors which may need to be accommodated or adjusted in some way:

Most countries will have a form of legislation designed to protect the interests of users with special needs or disabilities.

1) The time of the assessment
2) The place and physical properties of the examination environment
3) Properties/configuration of the assessment software
4) Properties of the client-side hardware which the examinee will be using.

With approximately 10% of males suffering from some type of colour-blindness making sure that colours do not combine in inappropriate ways should be a key design factor when writing exam questions. Colour can also influence the text perception of students with Scotopic Sensitivity Syndrome and larger font sizes may be required by students with other visual abnormalities. Ideally it should be possible to change the background colour of an online exam for anyone identified as being able to benefit from such colour changes.

Administration

- **Liaising with IT Services**: In parallel with room booking should be communication and agreement with the central institutional IT support unit. Keeping such a unit informed of timetabled summative assessments is vital so that planned maintenance of client computers, servers and networking infrastructure can be accommodated around the exam dates. In the UK the Joint Academic Network (JANET) that is used by all major universities has what is referred to as an 'at risk' period of 8am-10am on Tuesday mornings. Where possible online summative exams should not be scheduled during known at-risk times.

- **Starting the exam**: It is good practice to request that students report to the relevant computer lab 10-15 minutes ahead of the scheduled exam start time. This provides plenty of time to log into the system with their username/password (authenticate). Invigilators and IT support personnel should either have printed password lists or have access to a computer to look up the log in details of any student who forgets their details. It is also prudent for the assessment system administrators to create two or three temporary 'guest' accounts which can be given out to any unexpected students who need to sit the exam.

- **Disaster management**: There should be a faculty/departmental disaster recovery protocol document. This should ideally cover points from guidance sources such as BS ISO/IEC 23988 (2007) but be grounded in the specific practicalities of the assessment system used. For example, one of the most common disaster recovery activities is likely to be dealing with the crash of a single student's computer. In such circumstances the invigilators or IT support staff should be able to take steps to move the student to a spare computer and to restart the exam with as little loss of data as possible. Some systems require the user to explicitly save information; some save information automatically between screens and others save automatically at periodic intervals. Knowing the precise mechanisms used by the assessment system in use will allow the disaster recovery protocol document to be fine tuned. Another event which should be planned for is a fire evacuation in the middle of an examination. Systems such as TouchStone (2007) contain 'fire exit' icons which when pressed do two things:

  1) saves all data back to the server, and
  2) blanks the screen so that evacuating examinees cannot see the answers of their peers when leaving the lab.

Results Analysis

Moderation

With an assessment successfully delivered the results need to be analysed. The exact pass mark should be entered into the assessment software and the output reports should display a 'pass' or 'fail' descriptor next to each students' name. Most reports of this type will include broad statistical data such as maximum, minimum, mean and median scores for the cohort expressed as marks and percentages. These should be checked by the module coordinator or academic member of staff responsible for the assessment. In the United Kingdom this manual checking of the results is an important legal step as under the Data Protection Act (1998) there are clauses which provide rights to individuals that give protection against decisions based on personal data made solely automatically. It is advisable to discuss in more detail relevant legislation with a data protection officer at your institution.

Assuming the marks appear roughly in line with what is expected the marks will normally need to be transferred to some kind of student management information system. Each system will differ in the format of the required data,

however, the goal is to try and ensure an automatic transfer process. Most assessment systems will provide a variety of data outputs, the common being MS Excel, comma-separated values (CSV) files or XML files.

Item Analysis

Having considered how the examinees performed attention can be turned to how well the question items performed. There are a number of different forms of investigations which come under the umbrella term 'item analysis'. At this point the reader is directed to the summary provided by McAlpine (2002) covering the most common three: Classical Test Theory, Item Response Theory and Rasch Measurement. The range of available analyses will depend on the specific assessment system being used; however, many systems will support some sort of data export which may then be entered into a specific statistical package for further processing. The results analysis phase, although the last part of the summative assessment lifecycle, represents the first step of the coming academic year feeding into both future teaching plans and question writing.

Where an item is found to have performed poorly there should be agreed departmental policies for investigation. The first step is probably to check that the correct answer has been accurately set within the assessment system. If it has been incorrectly set then the question should be corrected and the students' answers remarked (this step might be automatic in some systems). Alternatively if the answer is correctly set on a poorly performing question then a number of actions may be taken: a) it could be removed from the paper and the students' responses remarked, b) the results of the analysis communicated back to the question author(s) so it may be amended in future, and c) changes to the curriculum made to explain concepts that are misunderstood by the majority of the cohort.

Where an item is found to have performed poorly there should be agreed departmental policies for investigation.

Future Trends in on-line eAssessment

Reduced time spent marking is probably the most often cited advantage of moving towards computer-based assessment, but it will be interesting to see how long it takes the market place to move from online assessment as merely delivery to an integrated part of the whole process.

Reduced time spent marking is probably the most often cited advantage of moving towards computer-based assessment.

Question Types

Systems such as TRIADS (2007) and Perception (2007) support many different question types that are not possible on paper but there is limited literature about the validity and reliability of these new forms. Intuitively the ability to drag and drop labels onto an image, for example, appears convincing but this needs to be studied scientifically. Research in this area will also be useful in encouraging more interactive question type use as it can be all too easy for the creation of online assessments to become a form filling exercise for simple multiple choice questions, rather than using these systems in ways that really sets them apart from examinations on paper. In addition to validity and reliability, research into how long it takes examinees to complete different question types would also make a useful contribution that should help question writers determine how long an exam should be.

Simulations

The use of simulations is likely to influence on-line eAssessments particularly if they are configured for server-client usage. The ability to assess how well a student interacts with a clinical or physiological model capable of undergoing a wide variety of pre-programmed changes creates an extremely powerful tool with high validity. Exposing learners to such situations is not only a powerful learning experience but can be an important way of assessing safety and competence.

Computer Adaptive Testing (CAT)

Adaptive testing involves building more 'intelligence' into the assessment system so it can monitor and interact with the user's input (Green 2000). Depending on how well a user answers questions the software can provide feedback and create an appropriate path through the assessment system designed to test the user's knowledge to a pre-determined limit. For example a correct answer might be followed by a more difficult question or an incorrect answer by an easier question (Schuwirth and van der Vleuten, 2006). An Item Response Theory (IRT) model can be used to monitor, evaluate and record overall activity and to provide feedback reports to learners and teachers (Rudner 1998).

Textual analysis

This Guide has concentrated on the on-line assessment of objective tests where, by the very nature of the system, there is little or no ambiguity concerning the responses learners input. This form of assessment excludes essays and short text based answers which inevitably blocks the range of assessment formats that learners might prefer to demonstrate their knowledge and understanding. Marking essays and short answers is also time consuming and potentially less reliable than objective testing. However, the technology to assess essays and short answers is slowly developing and it is likely that in the next few years systems will be developed that can perform these tasks satisfactorily. For a review of developments in the field see Valentini et al. 2003.

Management Challenges

One of the key challenges for the future of eAssessment will be moving from a cottage industry approach lead by individuals to a mass-produced system where quality is overseen by a management process. Many early innovators of eAssessment were lead by talented individuals with vision, drive and passion. Typically either an IT expert would push the capabilities of a new assessment system or an academic would push for early use of a system, replacing traditional assessment methods. However, as the benefits of eAssessment are repeatedly demonstrated and its use spreads to other modules, degrees and faculties, the problem of how to manage the whole process begins to grow. In the early stages of eAssessment adoption the individuals driving the change often take on multiple roles: training students how to use the system, writing the questions, being on hand during exams in case of problems, and so on. However, scaling up the endeavour requires clearer roles for a wider variety of stakeholders. Some stakeholders, such as exam boards, may change little in the move from paper-based to computer-

One of the key challenges for the future of eAssessment will be moving from a cottage industry approach lead by individuals to a mass-produced system where quality is overseen by a management process.

based assessment. Others, for example external examiners, may have to change a lot by being asked to log into the exam paper online with their own usernames/passwords and then asked to submit comments electronically. The precise changes in role will depend on two key factors, the institutional approach, and the eAssessment system employed.

The institutional approach is an important factor because some are creating specialist eAssessment units that take on the whole process (excluding question creation). Other institutions are working with a more distributed approach similar to paper-based exams whereby a large number of individuals all contribute to the assessment process by having clearly defined roles: question writer, time-tabler, external examiner, trainer, etc. It seems likely that more institutions will favour this distributed approach to roles, especially when some exams may be on paper and others computer-based.

The second factor that will influence the specific roles of various stakeholders is the capabilities of the eAssessment system adopted. For example, some systems support external examiner access and standard setting, whereas others do not. An audit must be made of which parts of the overall assessment lifecycle can be facilitated online and which cannot.

TABLE 1
Roles and Responsibilities

PHASE	ACADEMIC	ADMINISTRATIVE	INFORMATION TECHNOLOGY
Pre-exam	**Module Convenor**: Selects appropriate format for assessment **Question writer(s)**: Questions are written by academics, could include module convenor **Question imputer**: Questions require entry into e-assessment system. Sometimes this is cut and paste from Word, sometimes questions are written directly into e-assessment system. **Internal question reviewers**: Questions should be reviewed by experts in the subject matter for any problems. **Standards Setting team**: For subjects using standards setting techniques, a team meeting must be arranged to facilitate this process. **External question reviewers**: Access to the completed paper needs to be given to an external examiner. Ways should be found to facilitate this online.	**Exam scheduling**: A timetable of exams is created for each module. **Room Booking**: In conjunction with the exam timetabling must be the booking of computer labs with sufficient numbers of computers. **Accessibility Unit**: Examinees with special needs, such as dyslexia, must be identified and various adjustments made so that they are not unfairly disadvantages at exam time.	**Server support / security**: The server on which the e-assessment system resides requires regular maintenance and security updating. Needless to say this must be performed with knowledge of the exam timetable. **Networking**: Teams responsible for the local area network must be notified of exam times so that maintenance to routers/switches, etc can be planned. **Trainers for Staff**: Staff must be trained in: a) the capabilities of the e-assessment system, and b) how to create questions/papers in the system well in advance of the scheduled exam date. **Trainers for Students**: The students must be informed that they should expect some of their exams to be online and how the software works for each question type.
During exam	**Academic source for mid-exam problems**: As with paper exams an academic should be on hand in case of any content problems.	**Invigilators**: As with paper-based exams invigilators or proctors are required to reduce plagiarism.	**CAA software support**: Staff knowledgeable in the e-assessment system in use must be on hand during exam time in case of any crashes or other problems.
Post-exam	**Moderators**: After the exam is complete the performance of the exam cohort should be examined and any poorly performing questions removed/moderated. **Exam Board**: Final moderated marks should be sent to a formal exam board.		

The management challenge, as the adoption of eAssessment becomes wider across the sector, is the establishment of structures to ensure question quality, plus co-ordinated administrative and IT provision. The key to these new management structures will be clear definitions and, if necessary, repurposing, of individuals' roles. Failure to define roles will damage the eAssessment process and lead to confusion, reinforcing the divides between academics, administrators and IT support personnel. Clear responsibilities can build bridges between these groups and foster trust through mutual appreciation of each others' role.

Table 1 summarises the roles and responsibilities required of academics, administrators and IT staff during the exam cycle to ensure an effective eAssessment process.

Conclusions

It is the intention of this Guide to demonstrate how computer-based assessment can and should be integrated into the wider assessment process. As mentioned in the introduction there have been a few documented failures of high-profile summative examinations (Harwood, 2005; Heintz & Jemison, 2005; Smailes, 2002) and it is tempting to suggest that the commonality between them is IT failure. While it appears that it was hardware and network speed issues that lay behind the failures it is also likely that it was a failure to fully engage in the communications process between all parties that ultimately resulted in the cause of the failures.

One of the difficulties of the communication process that must be overcome is differences in the language used between stakeholders. Academic staff will use a certain vocabulary, such as pedagogy, curricular alignment and cognitive difficulty, administrators will use their terms such as cohort, session, entry year and so on, and IT staff will use terms such as load, performance and bandwidth. While the reader may think they are familiar with the terms listed here, making sure that all are understood and that the same meaning is attributed to them by all parties is vital. The terms 'reliability' and 'performance' will be used by both academics and IT specialists when referring to assessment, but the context and therefore the meaning of such terms are completely different.

It is hoped that the reader at this stage who is interested in trying to pilot the introduction of online summative assessment into his or her institution feels suitable informed to be able to start the process going. As just mentioned, this is a process that at its core is a communications exercise between a wide variety of different stakeholders. Those stakeholders must come together to create assessments that should be defendable intellectually, legally, technically and economically. Keeping these four perspectives in mind the chapter outlined some of the more important issues to be considered during each of the five stages of the assessment development lifecycle suggested by UKCDR (2007). Adopting the principles set out here should create an accountable and robust online assessment process that can withstand scrutiny.

References

ANGOFF WH (1971). Norms, scales, and equivalent scores. In R.L. Thorndike (Ed). *Educational measurement* (2nd ed.). Washington, DC: American Education on Education.

BOUD D (1995). *Enhancing learning through self assessment*. London: Kogan Page

BS ISO/IEC 23988 (2007). Code of practice for the use of information technology (IT) in the delivery of assessments.

CANTILLON P, IRISH B & SALES D (2004). Using computers for assessment in medicine. *British Medical Journal*, 329, 606-609.

CASE SM & SWANSON DB (2002). (3rd edition) *Constructing Written Test Questions For the Basic and Clinical Sciences*. National Board of Medical Examiners, Philadelphia.

CRONBACH LJ (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*. 16, 297-334.

DATA PROTECTION ACT (1998) Data Protection Act 1998, Retrieved June 5, 2007 from http://www.opsi.gov.uk/acts/acts1998/19980029.htm

DENT JA & HARDEN RM (2005). *A Practical Guide for Medical Teachers*. Elsevier Limited.

EBEL RL (1972). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice-Hall

GREEN BF (2000). System design and operation. In Wainer, H. (Ed.) *Computerized Adaptive Testing: A Primer*. Mahwah, NJ: Lawrence Erlbaum Associates.

HARPER R (2003). Correcting computer-based assessments for guessing. *Journal of Computer Assisted Learning*, 19(1).

HARWOOD I (2005). When summative computer-aided assessment go wrong: disaster recovery after a major failure. *British Journal of Educational Technology*, 36(4): 587-597.

HEINTZ J & JEMISON J (2005). Online Exams: Embrace the Opportunity, Avoid the Pitfalls. 2005 EDUCAUSE Annual Conference. Retrieved June 3, 2007 from http://connect.educause.edu/blog/podcaster/e2005_podcast_online_exams/2027?time=1165945815

HINMAN LM (2000). Approaches to Cheating and Plagiarism: 11/2/2000. http://www.academicintegrity.org/Values.asp. Accessed April 24th 2006.

HOLSGROVE G & ELZUBEIR M (1998). Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Medical Education, 32(4): 343-350.

LIU M, PAPATHANASIOU E & HAO Y (2001). Exploring the use of multimedia examination formats in undergraduate teaching: results from the fielding testing. *Computers in Human Behaviour*, 17: 225-248.

MCALPINE M (2002). *A Summary of Methods of Item Analysis*. CAA Centre, Luton.

MCKENDREE J (2006). *ASME Understanding Medical Education guides*. eLearning. Association for the Study of Medical Education. Edinburgh.

NIELSEN J (2005). Ten Usability Heuristics. Retrieved May 20, 2007 from http://www.useit.com/papers/heuristic/heuristic_list.html

NORCINI JJ (2003) The metric of medical education: setting standards on educational tests. *Medical Education*, 37: 464–469

OBLINGER D (2006). *Listening to what we're seeing*. Keynote Speech at the 2006 ALT-C conference, Edinburgh, United Kingdom.

PERCEPTION (2007). Question Mark. Retrieved May 3, 2007 from http://www.qmark.com/uk/index.aspx

RUDNER LM (1998). An On-line, Interactive, Computer Adaptive Testing Tutorial, http://edres.org/scripts/cat

SCHUWIRTH & VAN DER VLEUTEN (2006). *ASME Understanding Medical Education guides. How to design a useful test: the principles of assessment*. Association for the Study of Medical Education. Edinburgh.

SENDA (2001). Special Educational Needs and Disability Act 2001, Retrieved May 20, 2007 from http://www.opsi.gov.uk/acts/acts2001/20010010.htm

SIM G, HOLIFIELD P & BROWN M (2004). Implementation of computer assisted assessment: lessons from the literature. *ALT-J*, 12(3): 215-230.

SIM G, STRONG A & HOLIFIELD P (2005). *The design of multimedia assessment objects*. Proceedings for 9th CAA Conference, Loughborough, United Kingdom.

SIMMAN (http://www.laerdal.com/) (Retrieved 1/1/2008)

SMAILES J (2002). *Experiences of using Computer Aided Assessment within a Virtual Learning Environment*. BEST 2002.

TAYLOR WL (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly*, 30: 415-433.

TOUCHSTONE (2007). University of Nottingham Medical Education Unit. Retrieved May 3, 2007 from http://www.nottingham.ac.uk/nle/about/touchstone/

TRIADS (2007). Centre for Interactive Assessment Development. Retrieved May 3, 2007 from http://www.derby.ac.uk/ciad/

UKCDR (2007). UK Collaboration for a Digital Repository for High Stakes Assessments. Retrieved May 3, 2007 from http://www.ukcdr.manchester.ac.uk/

UNDERWOOD J (2006). *Digital Technologies and Dishonesty in Examinations and Tests*. Qualifications and Curriculum Authority.

VALENTINI SALVATORE, FRANCESCA NERI & ALESSANDRO CUCCHIARELLI (2003). An Overview of Current Research on Automated Essay Grading. *Journal of Information Technology Education*, 2: 319-330.

WOOD D (2007). *Formative Assessment*. Association for Medical Education. Edinburgh.

Additional Reading

ASSESSING LEARNING IN AUSTRALIAN UNIVERSITIES. Centre for the Study of Higher Education. 'On-line assessment', Retrieved 1/1/2008 from http://www.cshe.unimelb.edu.au/assessinglearning/03/online.html

BENYON D, CRERAR MA & WILKINSON S (2001). *Individual Differences and Inclusive Design*. User Interfaces For All: Concepts, Methods, and Tools. Lawrence Erlbaum Associates, Inc.

BULL J & DANSON M (2004). Computer-assisted Assessment (CAA). Learning and Teaching Support Network (LTSN). Retrieved June 14, 2007 from http://www.heacademy.ac.uk/embedded_object.asp?id=20388&prompt=yes&filename=ASS093

HOPKINS KD (1998). *Educational and Psychological Measurement and Evaluation*. Allyn & Bacon (8th edition).

JISC (2007). Effective Practice with e-Assessment: An overview of technologies, policies and practice in further and higher education. HEFCE. Retrieved June 14, 2007 from http://www.jisc.ac.uk/media/documents/themes/elearning/effprac_eassess.pdf

PROCEEDINGS FOR 10TH CAA CONFERENCE (2006).

SCOTTISH QUALIFICATIONS AUTHORITY (2003). SQA Guidelines on Online Assessment for Further Education. Scottish Qualifications Authority. Retrieved June 14, 2007 from http://www.sqa.org.uk/files_ccc/GuidelinesForOnlineAssessment(Web).pdf

UNIVERSITY OF TECHNOLOGY SYDNEY 2008). Institute for Interactive Media and Learning. 'On-line Assessment'. (Retrieved 1/1/2008) http://www.iml.uts.edu.au/assessment/online/

Web references for on-line eAssessment systems

There are a number of eAssessment systems available either commercially or as free-ware for non-commercial use. The following list of URLs provides access to information about the main types available:

CASTLE TOOLKIT (Retrieved 1/1/2008)
http://www.le.ac.uk/castle/

E-ASSESSOR (retrieved 1/1/08)
http://ferl.qia.org.uk/display.cfm?page=655

EXAM BUILDER (Retrieved 1/1/2008)
http://www.exambuilder.com/

HOT POTATOES (Retrieved 1/1/2008)
http://hotpot.uvic.ca/

NEWSLATE (Retrieved 1/1/2008)
http://www.newslate.com/

QUIA WEB (Retrieved 1/1/2008)
http://www.quia.com/web/

QUESTIONMARK PERCEPTION (Retrieved 1/1/2008)
http://www.questionmark.com/uk/index.aspx

QUESTION TOOLS (Retrieved 1/1/2008)
http://www.questiontools.com/index.html

QUIZ CENTER (Retrieved 1/1/2008)
http://school.discovery.com/quizcenter/quizcenter.html

Test Nation (Retrieved 1/1/2008)
http://www.testnation.net/

TOUCHSTONE (Retrieved 1/1/2008)
http://www.nottingham.ac.uk/nle/about/touchstone/

WEBASSIGN (Retrieved 1/1/2008)
http://www.webassign.net/

45 Portfolios for Assessment and Learning

Jan van Tartwijk, Erik W Driessen (2010)
ISBN: 978-1-903934-57-9

An overview of the content and structure of various types of portfolios, including eportfolios, and the factors that influence their success.

46 Student Selected Components

Simon C Riley (2010)
ISBN: 978-1-903934-58-6

An insight into the structure of an SSC programme and its various important component parts.

47 Using Rural and Remote Settings in the Undergraduate Medical Curriculum

Moira Maley, Paul Worley, John Dent (2010)
ISBN: 978-1-903934-59-3

A description of an RRME programme in action with a discussion of the potential benefits and issues relating to implementation.

48 Effective Small Group Learning

Sarah Edmunds, George Brown (2010)
ISBN: 978-1-903934-60-9

An overview of the use of small group methods in medicine and what makes them effective.

49 How to Measure the Quality of the OSCE: A Review of Metrics

Godfrey Pell, Richard Fuller, Matthew Homer, Trudie Roberts (2011)
ISBN: 978-1-903934-62-3

A review of the metrics that are available for measuring quality in assessment and indicating how a rounded picture of OSCE assessment quality may be constructed by using a variety of measures.

50 Simulation in Healthcare Education. Building a Simulation Programme: a Practical Guide

Kamran Khan, Serena Tolhurst-Cleaver, Sara White, William Simpson (2011)
ISBN: 978-1-903934-63-0

A very practical approach to designing, developing and organising a simulation programme.

51 Communication Skills: An essential component of medical curricula. Part I: Assessment of Clinical Communication

Anita Laidlaw, Jo Hart (2011)
ISBN: 978-1-903934-85-2

An overview of the essential steps to take in the development of assessing the competency of students' communication skills.

52 Situativity Theory: A perspective on how participants and the environment can interact

Steven J Durning, Anthony R Artino (2011)
ISBN: 978-1-903934-87-6

Clarification of the theory that our environment affects what we and our students learn.

53 Ethics and Law in the Medical Curriculum

Al Dowie, Anthea Martin (2011)
ISBN: 978-1-903934-89-0

An explanation of why and a practical description of how this important topic can be introduced into the undergraduate medical curriculum.

54 Post Examination Analysis of Objective Tests

Mohsen Tavakol, Reg Dennick (2012)
ISBN: 978-1-903934-91-3

A clear overview of the practical importance of analysing questions to ensure quality and fairness of the test.

55 Developing a Medical School: Expansion of Medical Student Capacity in New Locations

David Snadden, Joanna Bates, Philip Burns, Oscar Casiro, Richard B Hays, Dan Hunt, Angela Towle (2012)
ISBN: 978-1-903934-93-7

As many new medical schools are developed around the world, this guide draws upon the experience of seven experts to provide a very practical and logical approach to this often difficult subject.

56 Research in Medical Education. "The research compass": An introduction to research in medical education

Charlotte Ringsted, Brian Hodges, Albert Scherpbier (2012)
ISBN: 978-1-903934-95-1

An introductory guide giving a broad overview of the importance attached to research in medical education.

57 General overview of the theories used in assessment

Lambert WT Schuwirth, Cees PM van der Vleuten (2012)
ISBN: 978-1-903934-97-5

As assessment is modified to suit student learning, it is important that we understand the theories that underpin which method of assessment are chosen. This guide provides an insight into the essesntial theories used.

58 Self-Regulation Theory: Applications to medical education

John Sandars, Timothy J Cleary (2012)
ISBN: 978-1-903934-99-9

Self-regulation theory, as applied to medical education, describes the cyclical control of academic and clinical performance through several key processes that include goal-directed behaviour, use of specific strategies to attain goals, and the adaptation and modification to behaviours or strategies to optimise learning and performance.

59 How can Self-Determination Theory assist our understanding of the teaching and learning processes in medical education?

Olle ThJ ten Cate, Rashmi A Kusurkar, Geoffrey C Williams (2012)
ISBN: 978-1-908438-01-0

Self-Determination Theory (SDT) serves among the current major motivational theories in psychology but its applications in medical education are rare. This guide uncovers the potential of SDT to help understand many common processes in medical education.

60 Building bridges between theory and practice in medical education by using a design-based research approach

Diana HJM Dolmans, Dineke Tigelaar (2012)
ISBN: 978-1-908438-03-4

This guide describes how Design-Based Research (DBR) can help to bridge the gap between research and practice, by contributing towards theory testing and refinement on the one hand and improvement of educational practice on the other.

61 Integrating Professionalism into the Curriculum

Helen O'Sullivan, Walther van Mook, Ray Fewtrell, Val Wass (2012)
ISBN: 978-1-908438-05-8

Professionalism is now established an an important component of all medical curricula. This guide clearly explains the why and how of integrating professionalism into the curriculum and ways to overcome many of the obstacles encountered.

62 Theoretical Insights into Interprofessional Education

Sarah Heam, Deborah Craddock, Marilyn Hammick (2012)
ISBN: 978-1-908438-07-2

This Guide supports the need for theory in the practice of interprofessional education and highlights a range of theories that can be applied to interprofessional education. It discusses specifically the application of theories that support the social dimensions of interprofessional learning and teaching, choosing by way of illustration the theory of social capital, adult learning theory and a sociological perspective of interprofessional education.

63 Experiential Learning

Sarah Yardley, Pim Teunissen, Tim Dornan (2012)
ISBN: 978-1-908438-09-6

This Guide provides an overview of educational theory relevant to learning from experience. It considers experience gained in clinical workplaces from early medical student days through qualification to continuing professional development.

64 Control-Value Theory: Using achievement emotions to improve understanding of motivation, learning and performance in medical education

Anthony R Artino Jr, Eric Holmboe, Steven J Durning (2012)
ISBN: 978-1-908438-11-9

This Guide considers the emergent theoretical and empirical work on human emotion and how this work can inform the theory, research and practice of medical education.

65 Team-Based Learning

Dean Parmelee, Larry K. Michaelsen, Sandy Cook & Patricia D. Hudes (2012)
ISBN: 978-1-908438-13-3

In times when available resources to Medical Schools are in short supply and classes are increasing in size, this Guide demonstrates how Team-Based Learning can use the educational values and principles of small group learning to improve the learning in large groups.

About AMEE

What is AMEE?

AMEE is an association for all with an interest in medical and healthcare professions education, with members throughout the world. AMEE's interests span the continuum of education from undergraduate/basic training, through postgraduate/specialist training, to continuing professional development/continuing medical education.

- **Conferences:** Since 1973 AMEE has been organising an annual conference, held in a European city. The conference now attracts over 2300 participants from 80 countries.
- **Courses:** AMEE offers a series of courses at AMEE and other major medical education conferences relating to teaching, assessment, research and technology in medical education.
- **MedEdWorld**: AMEE's exciting new initiative has been established to help all concerned with medical education to keep up to date with developments in the field, to promote networking and sharing of ideas and resources between members and to promote collaborative learning between students and teachers internationally.
- **Medical Teacher:** AMEE produces a leading international journal, Medical Teacher, published 12 times a year, included in the membership fee for individual and student members.
- **Education Guides:** AMEE also produces a series of education guides on a range of topics, including Best Evidence Medical Education Guides reporting results of BEME Systematic Reviews in medical education.
- **Best Evidence Medical Education (BEME):** AMEE is a leading player in the BEME initiative which aims to create a culture of the use of best evidence in making decisions about teaching in medical and healthcare professions education.

See www.amee.org for more information.

If you would like more information about AMEE and its activities, please contact the AMEE Office:

Association for Medical Education in Europe (AMEE), Tay Park House, 484 Perth Road, Dundee DD2 1LR, UK

Tel: +44 (0)1382 381953; Fax: +44 (0)1382 381987; Email: amee@dundee.ac.uk

www.amee.org

ISBN: 978-1-903934-73-9

Scottish Charity No. SC 031618