

## Key Findings from recent literature on Computer-aided Assessment

Bill Warburton and Grainne Conole, University of Southampton

Address for correspondence: wiw@soton.ac.uk

*A perusal of the literature on Computer-aided Assessment (CAA) yields a rich crop of case studies and 'should-do's'. This paper gathers together the key findings and common themes found in a search of recent papers published on CAA implementation, including some projects under the Teaching and Learning Technology Programme (TLTP) and the Fund for the Development of Teaching and Learning (FDTL) projects. It is hoped that this review will provide a valuable snapshot of current practice in CAA across different subject disciplines and a distillation of some of the key commonalities.*

*Initial findings indicate that barriers to embedding CAA in institutions include a lack of commitment at a strategic level, and as well as barriers in terms of cultural or pedagogical issues, rather more than any technical shortcomings in the CAA system. Furthermore, computer-assisted assessment in general is often still something of an add-on, and is not yet fully integrated into other learning and teaching practices. Technical and cultural barriers to the adoption of CAA were found to vary between subject boundaries (CAA Centre, 2001). The enablers to successful embedding of CAA in institutions include institutional commitment together with provision of adequate funds, time and personnel (Bull, 2002), frequent communication between the various stakeholders (Raine, 1999), interoperability and the existing availability of a Virtual Learning Environment (VLE) (Stevenson et al., 2001).*

### Introduction

There is pressure in Higher Education (HE) for 'better' and more frequent assessment while resources generally are seen to be static or dwindling (TQEC Cooke Report, quoted by Dearing, 1997 p.206; Farrer, 2002; Laurillard, 2002) and automating assessment is seen by many as a way to control the assessment load in HE. However, assessment is generally acknowledged to be a complex issue and is also widely regarded as the most critical element of learning, so the stakes are very high (Brown et al., 1997; Bull and McKenna, 2001; Conole and Bull, 2002; Laurillard, 2002; McAlpine, 2002b). This paper offers a summary of recent findings put in the context of earlier research (Hart, 1998 pp.189-191), and attempts to summarise the findings of recent research literature in the field of computer-aided assessment (CAA) especially regarding its implementation and evaluation on a large scale. This review should also form the basis of the literature review for a PhD in the implementation and evaluation of CAA, and constructive comments would be gratefully received. Computer Based Testing (CBT) means tests that are delivered by computer, whilst the more inclusive term 'CAA' covers both CBT and responses written on paper that are scanned into a computer system (Stephens and Mascia, 1997 p.2). A more inclusive definition of CAA is that it:

*'... encompasses a range of activities, including the delivery, marking and analysis of all or part of the student assessment process using standalone or networked computers and associated technologies.'* (Conole and Bull, 2002 p.64)

### Methodology

A search was conducted in a large number of published resources from 1998 up to date (August 2003). Criteria for inclusion in the review were that at least one of the main topics in the research should be to do with the implementation or evaluation of CAA, and some paper have been included even though they were published a few years earlier than the 1998 limit because they shed light on the recent literature. Some research papers that met these criteria were left out because they had

been superseded, for instance by later work from the same writer, or because they appeared to be minor contributions to the literature in which case they were excluded for reasons of space. The keywords chosen for the search were 'CAA', 'CBA', 'computer-based assessment', 'computer-assisted assessment', 'on(-)line assessment', 'assessment'. Because managed learning environments include a CAA component, the following terms were also searched for: 'MLE', 'managed learning environment', and because much CAA testing is currently done using objective items (REFS) the following words were also included: 'multiple-choice', 'MRQ' 'objective test(s)', 'objective item(s)', 'objective testing', 'item bank(s)'. The implementation and evaluation of large-scale CAA systems is also at issue so the phrases 'implementation of Learning Technology (LT)' and 'evaluation of Learning Technology' were incorporated.

#### *Journals, Conference proceedings and electronic resources searched*

The following journals were searched: *Assessment & Evaluation in Higher Education* (19 articles found that are more or less relevant), *British Journal of Educational Technology* (BJET) (13 articles found that are more or less relevant, most of which were found in a special issue on CAA & online assessment [Vol. 33 No. 5, November 2002]), *Computers in Education* (three articles found that are more or less relevant), *Computers in Human Behavior* (one article found that is more or less relevant) and *Journal of the Association of Learning Technology* (ALT-J)- seven articles found that are more or less relevant. The following On-line journals were searched: *Australian Journal of Educational Technology* (AJET): 90 matches found of which three are more or less relevant and *The Journal of Technology, Learning, and Assessment* [www.jtla.org] (two more or less relevant articles). The following Citation indices were searched: *British Education Index* (BEI), *Current Index to Journals in Education*, ERIC (13 matches found, 8 more or less relevant), Education Index, Educational Technology Abstracts and Social Sciences Citation Index. The following Conference proceedings were searched: 3<sup>rd</sup> (1999) through to the 7<sup>th</sup> (2003) International CAA Conference proceedings [www.caaconference.com]- 146 articles found, all of which as may be expected are relevant to this review. These collections searched included *Dissertation Abstracts* which stores mainly American theses; ten matches were found, but none were particularly relevant and [www.theses.com] where British doctoral theses are kept; two matches were found but abstracts did not indicate much relevance to the research. The following UK Web sites searched, including 'Grey' resources: the *British Educational Research Association's* (BERA) website [www.bera.ac.uk]- 34 matches found, one more or less relevant, the *Bath Information and Data Services* (BIDS) site [www.bids.ac.uk]- 41 matches found, 21 more or less relevant), the *CAA Conference* site at Loughborough University [www.caacentre.com]- three articles found that are more or less relevant, the UK Government *Department for Education and Skills* (DFES) site [www.dfes.gov.uk]- 20 matches found, none of which were particularly relevant], the *Joint Information Services Committee* (JISC) site [www.jisc.ac.uk]- more than 300 matches found, of which 12 were found to be more or less relevant, [www.ltsn.ac.uk]- six matches found, two more or less relevant, [www.scrolla.ac.uk]- 158 matches found, three more or less relevant, [www.scaan.ac.uk]- two matches found, one more or less relevant and [www.sosig.ac.uk]- 68 matches found, two of which were more or less relevant. Offshore Web sites searched included [www.aera.net]- no matches found, [www.asciilite.org.au]- 7 matches found, 4 more or less relevant and [www.tltgroup.org]- none found relevant.

#### **Key Findings**

This review comprises three sections which are intended to show a progression from issues of the most general kind (Section I) to the most specific (Section III):

- I. The theory underpinning CAA
- II. Strategic issues: - embedding CAA within the institution
- III. Evaluation of large scale LT projects

## I. The theory underpinning CAA

The theory of assessment in general is widely applicable to CAA, and many of the debates in CAA revolve around central issues of what constitutes good assessment practice, particularly regarding objective tests because so much CAA is based upon them. Items are learning objects that include all the components required to deliver a single objective question, and item banks are subject-specific collections of items. The details of item design and the limitations of existing item types are fertile areas for research in CAA; a major concern related to the nature of objective tests in the literature is whether multiple choice questions (MCQs) are really suitable for assessing Higher-order Learning Outcomes (HLOs) in HE students. Much research has been published on this and related questions (e.g. Davies, 2001; Farthing and McPhee, 1999; Ricketts and Wilks, 2001; Duke-Williams and King, 2001), and some of the positions taken could be summarised as ‘item-based testing is (1) inappropriate for examining HE students in *any* case (NCFOT, 1998), (2) appropriate for examining *lower-order learning outcomes* (LLOs) undergraduates (Farthing and McPhee, 1999) or (3) appropriate for examining *HLOs* in undergraduates, *provided* sufficient care is taken in their construction’ (Duke-Williams and King, 2001 p.12; Boyle et al., 2002 p.279).

MCQs and multiple response questions (MRQ) are the most frequently used (Bull, 1999; Stephens and Mascia, 1997; Warburton and Conole, 2003), but there is some pressure for the use of ‘more sophisticated’ question types (Davies, 2001). The classification of question types is thought by some to be an unresolved issue caused by marketing pressures that tempt vendors to inflate the number of question types supported by their CAA systems (Paterson, 2002).

Sclater’s introduction to Herd and Clark’s report on CAA in FE positions item banks as *the* crucial driver of CAA: ‘What will really make CAA work though is the development of large assessment item banks’ (Herd and Clark, 2003 p.2) whilst Mhairi McAlpine’s Bluepaper on question banks argues for the routine adoption of item banks based upon the vulnerability of CAA tests to challenges from students on the grounds of fairness, validity, security or quality assurance (McAlpine, 2002a p.4). The Electrical and Electronic Engineering Assessment Network (e3an) is a HEFCE Fund for Development in Teaching & Learning (FDTL) Phase 3 funded project that produced a public bank of 1400 items on Electrical and Electronic Engineering and has provided an understanding of cultural and subject-specific issues whilst producing guidelines and templates to support other similar initiatives (Bull et al., 2002; White and Davis, 2000).

Six kinds of examinable learning objective are distinguishable according to Bloom’s classic taxonomy (Bloom et al., 1956 p.18): simple regurgitation of knowledge is at the most fundamental level, rising through comprehension, application, analysis, synthesis and ultimately to evaluation. Other writers have suggested adjustments to it (Anderson et al., 2001; Krathwohl et al., 1964) or proposed their own alternative taxonomies (Fink, 2003), or proposed alternative agendas for the assessment of undergraduates, as in Bennett’s multiple levels of Critical Thinking (Barnett, 1997).

Some researchers differentiate simply between formative assessment that is primarily intended to facilitate learning, and summative assessment that is principally meant to assess students on what they have learnt (e.g. McAlpine, 2002b p.6; McAlpine and Higgison, 2001 p.4-4); some writers also consider diagnostic and self-assessment applications (e.g. Bull and McKenna, 2001 p.6; O’Reilly and Morgan, 1999 p.152-153). Sclater and Howie go further in distinguishing six different applications of CAA: ‘credit bearing’ or high-stakes summative tests that may be formal examinations or continuous assessment, formative self assessment that can be ‘authenticated’ self-assessment or anonymous self-assessment, and finally *diagnostic* tests that evaluate the student’s knowledge by pretesting before the course is commenced or post-testing to assess the effectiveness of the teaching (Sclater and Howie, 2003 p.285-286).

In addition to such papers that are largely the output of CAA researchers and practitioners, the CAA Centre (a Joint Information Services Committee (JISC)-funded Teaching and learning Technology Project (TLTP) strand 3 project that ran from 1999 to 2001) have produced a number of specialist

publications on CAA, including the 'Blueprint for CAA' (Bull and McKenna, 2001) and three Bluepapers written by Mhairi McAlpine from the perspective of the CAA community, the first of which outlines the basic principles of assessment (McAlpine, 2002b) and presents the basic concepts applicable to any rigorous assessment. The second deals with (objective) item analysis and covers the elements of Classical Test Theory (CTT) (2002b pp.3-12), the three basic models in Item Response Theory (IRT) (2002b pp.13-20) and a brief description of Rasch Measurement (2002b pp.21-25). CCT and IRT are presented uncritically, although she finishes by portraying some of the relative strengths and weakness of Rasch Measurement (2002b pp.24-25). Boyle et al. (2002) show all three modes of analysis- CTT, IRT and Rasch analysis- with a set of 25 questions used with 350 test-takers; they conclude that what they see as the present approach by many practitioners to CAA of neglecting the rigorous quality assurance (QA) of items is untenable, and this is presumably especially the case for medium and high stakes use- see Shepherd (2001), below. Boyle and O'Hare recommend that training in item construction and analysis should be obligatory for staff who are involved in developing CAA tests and that items should be peer-reviewed and trialled before use (Boyle and O'Hare, 2003 p.77).

Assessments may be 'high stakes', 'low stakes', or somewhere in between, and much of the pressure on academic and support staff who are running CAA tests derives from the influence that the outcome has on candidates' futures (Boyle et al., 2002 p.272). Shepherd's (2001) description of the properties attributable to low, medium and high stakes testing is summarised below:

| <i>Stakes</i>                       | <b>Low</b>     | <b>Medium</b>           | <b>High</b>          |
|-------------------------------------|----------------|-------------------------|----------------------|
| <i>Decisions:</i>                   | None           | Can be reversed         | Difficult to reverse |
| <i>ID individual:</i>               | None           | Important               | Very important       |
| <i>Proctoring:</i>                  | None           | Yes                     | Constant             |
| <i>Options:</i>                     | Study more     | Pass, fail, work harder | Pass or fail         |
| <i>Item &amp; test development:</i> | Minor          | Takes time              | Significant          |
| <i>Items created by:</i>            | Subject expert | Subject expert          | Subject expert       |
| <i>Statistics checked:</i>          | Subject expert | Time to time            | Psychometrician      |

**Table 1: Shepherd's (2001) interpretation of assessment stakes**

Computer-adaptive Testing (CAT) involves setting questions of difficulty that depends on the test-taker's previous responses. If a question is answered correctly, the estimate of his/her ability is raised and a more difficult question is presented and *vice versa*, giving the potential to test a wide range of student ability very concisely. Lilley and Barker (2003) constructed a database of 119 peer-reviewed items (the minimum is 100) and gave both 'traditional (non-adaptive) and CAT versions to 133 students on a programming course from the same item bank. Their approach to the adaptive element of assessment of is based on the three parameter IRT model where the probability of a student answering an item correctly is given by an expression that takes account of its difficulty, discrimination and guess factor. Lilley and Barker found that the students' results from the CAT version of the test correlated well with their results from the traditional version and that they didn't find the CAT test less fair (2003 p.179). They assert that because CAT items are written specifically to test particular levels of ability rather than all, it has the potential to deliver results that are more accurate and reliable than traditional CAA tests (2003 p.180).

Concerns regarding the possibility of naïve test-takers achieving passing scores in objective tests are addressed in two ways in the literature: first by discounting a test's guess factor- the 'mean uneducated guesser's score' (MUGS) of McCabe and Barrett (McAlpine and Hesketh, 2003; McCabe and Barrett, 2003), and secondly by adjusting the marking scheme away from simple tariffs where 'one correct answer equals one mark' to include the possibility of negative marking where incorrect responses are punished by being awarded negative scores. Confidence-based Assessment is where marks awarded for a response are predicted on a student's confidence that the correct response has been given (Davies, 2002; Gardner-Medwin and Gahan, 2003; McAlpine and Hesketh, 2003; McCabe and Barrett, 2003; Walker and Thompson, 2001).

## II. *Strategic issues: embedding CAA within the institution*

### *Generic issues in implementing any Learning Technology in Higher Education*

The Tavistock Institute (Sommerlad et al., 1999) evaluated a number of LT projects and found four basic implementation strategies that appeared to be successful in achieving institutional change, namely

- Negotiating entry and pitching the innovation at an appropriate level,
- Securing institutional support and getting the right stakeholders onside,
- Mobilising & engaging teaching staff and other key actors
- Diffusing technology-based teaching and learning innovations

The Association for Learning Technology (ALT) identified six ways that the strategic application of a learning technology such as CAA may add value to the efficiency and effectiveness of the learning process, and six other factors that may influence adversely the value it can add (ALT, 2003 p.6). These issues are conspicuous in much of the research on implementing CAA strategically.

ALT's potential benefits are:

1. *Opportunities to improve and expand on the scope and outreach of the learning opportunities they can offer students;*
2. *Ways to ensure equality of opportunity for all learners;*
3. *Alternative ways of enabling learners from cultural and social minorities, learners with disabilities, and learners with language and other difficulties to meet learning outcomes and demonstrate that they have been achieved;*
4. *Quality control and quality enhancement mechanisms;*
5. *Ubiquitous access opportunities for learners;*
6. *Enhanced opportunities for collaboration which may increase the re-usability of learning objects and resources.*

The potential pitfalls identified by ALT:

1. *The immaturity and volatility of some learning technology mean that there is a lot of work involved in keeping up with available products, especially with a market that is shaking out. Accordingly, much effort is wasted through poor understanding of the technology and its application.*
2. *There are a lot of products and services which are not especially suited to UK FE and HE pedagogic models.*
3. *It is possible to make expensive errors when there is a misalignment between technology, pedagogy and institutional infrastructure or culture. These errors are often repeated in parallel between educational institutions.*
4. *Standards and specifications are evolving, hard to understand, easy to fall foul of, and tend to be embraced with zeal, without the cost and quality implications being properly understood.*
5. *Much effort is also dissipated through a poor understanding of the theory and pedagogy that underpins the use of the technology.*
6. *The absence of a widely established and practiced methodology by which rigorously to evaluate e-learning, and through which to develop the secure body of knowledge on which to build learning technology as a discipline.*

### *What differentiates CAA from other Learning Technologies?*

CAA is differentiated from other LTs by the sensitive nature of high-stakes assessment generally; students are perceived to be increasingly litigious (QAA, 1998) and the clear scoring schemes of objective tests open the results of CAA tests to scrutiny and make any deficits in practice apparent in ways to which more traditional forms of assessment are not so susceptible. These considerations make thorough risk analysis and management strategies particularly important, as Zakrzewski and Steven point out in their suggested CAA implementation strategy (Zakrzewski and Steven, 2000).

Sclater and Howie described the characteristics of an 'ideal' CAA system in their (2003) study of user requirements. They point out that no contemporary 'off the shelf' commercial CAA system has been found to fulfil all of a HE institution's needs, and identify 21 possible roles that a full-scale CAA system might require that are gathered in six functionally groups as authors, viewers and validators of questions, test authors, viewers and validators, three roles associated with the sale of items or assessments, three administrative roles, six roles associated with processing the results of assessments and three to do the individual instances of tests- the test-taker, the timetabler and the invigilator (Sclater and Howie, 2003 p.286-289). McKenzie et al. recommend that CAA should:

*'deliver summative assessment in a reliable and secure manner, deliver informative and detailed feedback to the candidate, provide informative feedback on student performance to the tutor, provide the widest possible range of question styles so that assessment design is not compromised by the limitations of the system. provide a wide variety of question selection and sequencing options to facilitate deployment of assessments in a wide variety of delivery environments and modes of operation, have the ability to deliver assessments on stand-alone machines, over local area networks and across intranets and the internet via a web browser, provide sufficient data about questions to allow reliable item analysis to feedback into quality improvements in subsequent runs & provide an assessment development interface that facilitates the rapid and easy production of computer-based assessments with all the characteristics listed above'* (McKenzie et al., 2002 p.207-8)

### *Different scales of CAA implementation*

CAA is often implemented by individuals or small groups on *ad hoc* basis with no overarching strategy or IT infrastructure, which is thought by some to delay or even exclude its uptake on an institutional basis, or 'embedding' (Boyle and O'Hare, 2003 p.74; McKenna and Bull, 2000 p.27): the risks of such small-scale development are likely to include the usual generic risks of 'ploughing a lone furrow' such as isolation and underfunding although there are benefits to be had such as being able to get things done when one wants to (Kennedy, 1998). Some HE institutions are implementing CAA on a larger scale, and are encountering risks and benefits of a different order (Cosmans et al., 2002; Danson et al., 2001; Raine, 1999; Stevenson et al., 2002). The literature contains many example of small-scale *ad hoc* implementation which could be loosely grouped as formative, mixed and summative usage in contrast to full-scale use which covers every possible use. The following three recent cases exemplify each use and illustrate many of the concerns raised by ALT in its (2003) list of generic LT benefits and pitfalls.

#### *Small-scale CAA*

Wood and Burrow used the TRIADS CAA system in their (2002) *formative* assessments in Engineering in which they aimed to encourage a deep approach to learning and high pedagogic standards even though staff-student ratios were less favourable. In order to do this they required a CAA system that supported the assessment of HLOs with a wide range of question styles and that was capable of supplying directive formative feedback (2002 p.369). Half of the 38 students who completed the assessment returned the questionnaire. They generally found the questions easy to understand and the instructions to be clear; most wanted more feedback, especially with the Chemical Engineering assessment, and most found formative use of the CAA system helped them

assess their progress. Student feedback was mainly constructive and showed that formative assessment by appropriate CAA tools is valuable in encouraging independent student learning (2002 p.373). The Highway Drainage formative CAA tests were designed to test HLOs across the entire syllabus, and was made continuously available using web CD ROM distribution. More than 20 questions were developed with different question types to test as many as possible of the various levels of student learning and understanding (Anderson et al., 2001) although the metacognitive class of knowledge was excluded as non-subject-specific, although it was no question that tested Anderson et al's creative ability were written (2002 p.375-6). The Chemical Engineering questions took an average 2 hours preparation, and the Civil Engineering questions took 2½ hours each. The difference in time was attributed to the extensive feedback required for the Highway Drainage course, and unquantified periods of extra time was spent on peer review of questions and feedback (2002 p.378).

In McKenzie et al.'s (2002) *mixed* usage implementation, most tutors used the TRIADS CAA system for formative testing although some moved on to summative use. The pilot group consisted largely of academic tutors who were not trained in programming skills; some found the system complex and some 'fell by the wayside' (McKenzie et al., 2002 p.206). Those who persevered appreciated: 'the wide variety of question styles, its readily customized and flexible nature, the variety of feedback options, question styles that are suitable for honours-level students & the ease of handling questions requiring complex graphics' (2002 p.207). The authors emphasize that cultural barriers and insufficient time are by far the most significant impediments to the large-scale uptake of CAA (2002 p.207), and that many of the barriers to the uptake of CAA identified in earlier studies (Bull, 1999; Stephens and Mascia, 1997) are still apparent:

*'Unfortunately, many of those barriers still exist and it may require a substantial staff development push and re-organization of academic and support staff roles in order to realize the potential of computer-based assessment on a wider scale. Furthermore, whilst the functionality of assessment systems is limited to a relatively few, simple question styles then there will be scepticism amongst many academics that computer-based assessments are a valid form of testing in higher education'* (2002 p.207).

Kennedy in his (1998) small-scale home-grown, *summative* implementation of CAA describes the pressures for change in assessment (increasing student numbers, increasing demand for assessment, shrinking resources and wider geographical dispersion of students) which resulted in the implementation of a home-grown Unix-based CAA system that students accessed through a web browser (WEBMAT). 350 students had to be assessed and feedback given as quickly as possible. Some problems were noted: there were insufficient machines to test all the students concurrently; the network system overloaded, and some students had to wait 20 minutes to obtain a test page; the login ID of one student in 20 was rejected because they were taking an 'atypical' course; there were security concerns in that students could dump question screens to disc. In an institutional implementation, these issues are likely to have emerged at a relatively early stage (Zakrzewski and Steven, 2000); on the other hand, some of the benefits of using a home-grown system are clear- the solutions to many of the problems encountered, many of which are common to users of other CAA systems, were expedited by the author's intimate knowledge of the system and his ability to make changes to the underlying server. This may not have been so easy had he been using an 'off the shelf' package.

Best's small-scale summative project evaluated the use of Blackboard's quiz engine to assess was impeded by staff resistance and staff development opportunities not to be fully exploited; staff were busy and did not see CAA as a priority (Best, 2002 p.30). Her use of Blackboard led her to conclude that CAA must be designed to limit the amount of scrolling needed' (Best, 2002 p.34). Boyle et al's (2002) study of summative medium-stakes CAA tests in Mineralogy combined scores from objective and practical tests in a multi-modal assessment strategy that provided a more reliable overall measurement of student attainment than either alone, even though there was only a low

correlation between the two components themselves (2002 p.59). Assessments based solely on CAA should perhaps be viewed with caution, and care should be taken when relating the scores they give to scores from necessarily more subjective assessment methods. Although MCQs were found easy to implement and able to deliver reliable assessment, carry over negative marking produced relatively low student scores causing negative feedback. Boyle et al. recommend MRQs and internal negative marking as a more attractive alternative than ordinary MCQs (2002 p.59).

### *Large-scale CAA*

Much of the published research on the large scale implementation of CAA centres on the paramount importance of risk management, meaning that care must be taken to identify and classify the potential risks and then that realistic plans about how to avoid them must be put in place (Bull and McKenna, 2001 p.67; Zakrzewski and Steven, 2000). Zakrzewski and Steven describe an application of Boehm's (1988) spiral process model for software development to the implementation of CAA across an entire institution. Their adaptation of Boehm's model provides five phases of a sequence (Planning, Risk Analysis, Assessment Design, Evolutionary Development of System & Evaluation) that they say a large-scale CAA project should move through, the entire sequence being traversed recursively as the project moves from pilot stage through to full institutional roll-out. Thus the planning phase of each successive stage is based upon an evaluation of the previous stage, and the outcomes of the risk analysis and consequent risk management plan are addressed in the detailed design of the assessments. They assert that successful implementations of CAA depend on the risks being made explicit and thoroughly understood (Zakrzewski and Steven, 2000 pp.202-203) and proceed to break the perceived risks down into pedagogic, operational, technical, and financial kinds, many of which resonate with those identified in national surveys of CAA usage (Bull, 1999; Bull and McKenna, 2000; Bull, 2001; Stephens and Mascia, 1997; Warburton and Conole, 2003). Each individual risk is named and assigned weight during a 'brain-storming' process, and then all the risks are prioritised and appropriate solutions for their management devised and adopted. Some examples of specific risks that were identified and managed in this way at Luton are given (2000 p.206-211). The writers conclude that their model is applicable across any assessment infrastructure and that the only thing that should change for other institutions are the risks drawn from the model (2000 p.213). The following four implementations of CAA are on a large scale and illustrate many of these points and also those raised by ALT (ALT, 2003).

O'Leary and Cooke's Bristol's site-wide VLE (Blackboard) met many of the institution's CAA needs, but did not address issues of portability, robustness and security (O'Leary and Cook, 2001 p.9). A cross-institution staff survey revealed that lack of time, training, infrastructure (workstations and support) were all barriers to the uptake of CAA, but that increases in student numbers was perceived to be a driver for the uptake of CAA. At Bristol, different departments were already using different CAA systems, and an evaluative project (VIOLET) was established to test a number of existing alternative CAA solutions in order to determine which approach would best meet the stated goal of establishing a customised V/MLE that might also provide integrate with other coexisting CAA systems or VLEs (2001 p.5). Key concerns about CAA systems that arose from focus groups and interviews conducted by the VIOLET team included portability of learning objects including CAA tests and items, usability, width of feature set, Quality Assurance of published learning objects and technical/security issues. Subject-specific differences made it difficult to adopt any one system. Findings were that established and disparate practice made the continued co-existence of disparate CAA systems attractive, and that a way of integrating all the university's systems would have to be found, although it was recognised that a single future 'off the shelf' system might emerge that leap-frogged all these concerns (2001).

In an institution-wide implementation of Perception at a Belgian university, it was found that technical restrictions in Perception 2.5 resulted in the institution building its own semi-automated solution for uploading assessments. Gatekeeper releases assessments to preview server and after preview period, to delivery server. Manual intervention was required for republishing. The CAA



system accommodated the increased use of formative assessment necessitated by the institution's move to semester-based assessment where exams must be passed at the end of each semester. Customisation required for site-wide access (Cosemans et al., 2002).

At Loughborough University, the Teaching Quality Enhancement Fund (TQEF) funded the first part of a CAA/MLE project comprising a central web-based Perception (2.5) CAA system and the second part is funded by the Joint Information Systems Committee (JISC) Committee for Integrated Environments for Learners (JCIEL), and aimed to integrate the central web based assessment system with management information systems (MIS) thus moving towards a Managed Learning Environment (MLE); this part is not described in detail, although the one-way population of the Perception security database by Perl script from the MIS system is mentioned. Technical restrictions in Perception 2.5 resulted in the institution building its own solution for uploading assessments. Danson et al. note the paramount importance of understanding the institution's existing assessment procedures before beginning a CAA implementation:

*“A full picture of the assessment process should be taken into account when considering how the new CAA system will impact on the University”* (Danson et al., 2001 p.5)

A Gatekeeper publishes assessments to preview server, as with other large-scale implementations of Perception 2.5, and manual intervention is required to publish assessments for delivery. Extensive customisation was required to implement site-wide access, and a mass staff development scheme was planned for when the project finished (Danson et al., 2001).

Increasing student numbers and declining resources led Stevenson et al. to implement *WebCT* on a large scale at the University of Ulster with the aim of increasing the frequency of assessment without increasing tutors' workload. They created 600 items for formative and summative testing, and the results of CAA tests were compared with those from earlier traditional paper-based tests. Findings were that the CAA tests results were largely indistinguishable from the paper-based tests (Stevenson et al., 2002 p.334), that the use of question banks in general and automated items of the WebCT 'calculated question' type in particular is effective in reducing the incidence of cheating, and that the robustness of the system stemmed largely from the strength of institutional commitment (2002 p.333).

#### *Compliance with published standards for CAA practice*

The recent *Code of practice for the use of information technology (IT) in the delivery of assessments*- better known as BS 7988: 2002- acknowledges that increased use of CAA:

*‘has raised issues about the security and fairness of IT-delivered assessments, as well as resulting in a wide range of different practices’* (BSI, 2002 p.ii).

BS 7988:2002 aims to enhance the status of CAA and encourage its wider use in appropriate applications by demonstrating its fairness, security, authenticity and validity (2002 p.ii). However, the *Code of Practice's* focus on the delivery of CAA tests could lead to the relative neglect of earlier stages in the preparation and quality assurance of assessments:

*‘A poor assessment, delivered appropriately, would conform to BS 7988’* (Boyle and O'Hare, 2003 p.72).

Boyle and O'Hare identify the American Educational Research Association (AERA) *Standards for Educational and Psychological Testing* (2003 p. 71), the Association of Test Publishers (ATP) *Guidelines for Computer-based Testing* and the Scottish Qualifications Authority (SQA) *Guidelines for Online Assessment in Further Education* (2003 pp.71-72) as guides to good CAA practice in as far as they encourage a more all-round treatment of the entire production process including the selection of appropriate question types and analysis of item quality (SQA, 2003 pp.17-21).

*Surveys of CAA use: Barriers & Enablers to the uptake of CAA*

Three national surveys have been conducted on the use of CAA in HE (Bull and Hesketh, 2001; Stephens and Mascia, 1997; Warburton and Conole, 2003), and some surveys of internal practitioner opinion have conducted as part of large scale evaluations of CAA (e.g. Herd and Clark, 2003). These surveys shed light on the drivers, enablers and barriers to the uptake of CAA in HE (or Further Education (FE) in the case of Herd & Clark).

The greatest institutional barrier to the uptake of CAA was seen to be cost both in terms of personal time invested and the expense of commercial 'shrink-wrapped' CAA software (Bull, 1999 p.6; Warburton and Conole, 2003 p.437). Unrealistic expectations that may stem from a naivety regarding the underlying theory and practice of CAA, coupled with inherent conservatism and lack of technical and pedagogic support. Respondents were less concerned with MLE integration or, security and copyright issues (Warburton and Conole, 2003 p.437). The greatest obstacle to CAA uptake by individual academics was perceived to be lack of time exacerbated by the perceived steep learning curve associated with getting to grips with the technology and constructing specialized CAA question types, including the difficulty of constructing objective items that reliably assess HLOs which resonates with the (2003) findings of Boyle and O'Hare (ibid.). There was a perceived credibility gap between what CAA proponents promise and what respondents thought could reasonably be delivered; lack of support, cultural resistance to change and technophobia were cited less often. Related issues of unfriendly software, academics working in isolation and individual inertia were also raised (ibid.). Subject-specific shared question banks (Herd and Clark, 2003 p.21) and exemplars were cited as important drivers for the large-scale uptake of CAA, but interestingly, the provision of CAA 'evangelists' and adherence to institutional guidelines was thought less crucial. At an individual level, academic commitment and overcoming initial user barriers were cited as important enablers to the uptake of CAA; faculty support for CAA seems to be limited (mainly restricted to occasional time release) and it appears that external funding is the principle way that support for CAA at this level is rendered. Other important factors cited included the need to embed CAA within normal teaching, issues of effective interoperability (particularly between CAA systems and VLEs), integration of multimedia and reliable confidence testing within CAA systems (Warburton and Conole, 2003 p.438).

Lay and Sclater identify two reasons why the interoperability of question items and tests may be seen to be important in embedding CAA: will the question banks being created today be accessible in future years when the current CAA systems are no longer in use, and can student assessment data be transferred from the CAA system to the institutional student records system (Lay and Sclater, 2001 p.1). Another obvious and important driver for interoperability is preserving users' investments in existing questions and tests when moving between institutions or to a different CAA system. The IMS Consortium's Question and Test Interoperability (QTI) specification (IMS, 2003) is the leading contended for a CAA *lingua franca* (Lay and Sclater, 2001; Sclater et al., 2002) and the JISC-funded *Tools for Online Assessment Interoperability* (TOIA) project will produce a QTI-compliant full-featured CAA system that will be free for academic use in HE and FE (TOIA, 2003).

The most recent survey asked about the delivery modes for CAA and most was web-based, although a large fraction of respondents delivered CAA using closed networks, and a small percentage used OMR. Only a third of CAA tests were invigilated, and most of the summative CAA tests restricted the percentage weighting to a third or less, although it was noteworthy that in a small number of cases the CAA test was worth 100% of all marks awarded for a module. CAA was used test both small and large (greater than 200) groups of students (Warburton and Conole, 2003 p.438).

### III. *Evaluating the use of Learning Technology generally and CAA in particular*

The Learning Technology Dissemination Initiative (LTDI) produced a 'cookbook' for the evaluation of LT that encourages evaluators to ask basic questions about evaluation before embarking upon such an exercise, warning that-

*'Evaluations are costly. Even the simplest takes precious time from other activities.'* (Harvey, 1998 p.8)

The *Cookbook* suggests that close attention should be paid at a preliminary stage to identifying the stakeholders, understanding what the evaluation is designed to achieve, what kind of scale the evaluation will be on, its likely impact upon the institution and what will be delivered in terms of output- an internal report, a published paper, a website etc. (Harvey, 1998 pp.8-11). It proceeds to describe criteria for choosing appropriate methodologies for evaluating LT and then summarises a number of popular methodologies thought suitable for evaluating LT including 'traditional' evaluation tools such as questionnaires and checklists but adding some less traditional tools as well- user confidence logs, focus groups, resource questionnaires and system log data are examples (Harvey, 1998 pp.26-62).

Bull and McKenna argue that evaluation of CAA practice is 'a neglected activity' (Bull and McKenna, 2001 p.80) and agree with the writers of the *Cookbook* in suggesting that when preparing to evaluate an implementation of CAA, a number of factors must be considered first: attempts to assess the impact of a CAA system should be grounded in evaluation of learning technology in general; the purposes of the evaluation should be identified from the perspective of stakeholders; an appropriate evaluation methodology should be chosen from the alternatives to ensure a good fit with the project itself (Bull and McKenna, 2001 p.81). A number of tools for evaluating the use of technology in education have been proposed in recent years, including Laurillard's conversational framework (Laurillard, 2002), Conole and Oliver's pedagogical framework (Conole and Oliver, 1999) and Beetham's (2001) institutional audit tools (Beetham, 2001).

The search also found three evaluations of recent large-scale applications of LT in HE: Parker's (2001) evaluation of the TELRI project, Harvey & Oliver's (2001) final report on the EFFECTS project and Davis and Oliver's (2001) final report on the ASTER Project. An analysis of all three was undertaken in order to identify evaluation strategies that might be suitable for full-scale CAA.

#### *The TELRI project*

The HEFCE-funded TLTP3 TELRI project aimed to encourage the use of LT in research-led institutions to improve the quality of learning, especially the attainment of HLOs (Parker, 2001 pp.5-6). The report's methodology is 'an evaluative overview' of the project in two old research-led universities, one of which is relatively new and the other amongst the oldest (Parker, 2001 p.19). Data was gathered from by survey using a written questionnaire (sent to 264 users with a low return rate of 9%) that was in turn based upon an earlier interim survey by semi-structured interview of members of the project's steering group (Parker, 2001 p.5). Parker reported that the TELRI project team 'experienced varying degrees of enthusiasm and success in terms of their relationships with individual departments' (Parker, 2001 p.4) which appears to be a coded reference to the apparent resistance of some departments to change of any kind. This opposition was found to be so entrenched that the project was refocused at the end of the first Phase into two separate tracks for the second phase: 'Round 1' ('fast track') departments that were willing and able to make progress with LT, and 'Round 2' departments that required more basic help with clarifying their own learning objectives and assessment criteria (2001 pp.9-10). Many of the pitfalls identified by ALT (ALT, 2003 p.6) crop up in different words (Parker, 2001 p.11), and some of the greatest barriers to embedding LT in the TELRI project were found to be cultural, reflecting once again the finding of many others when assessing attempts to embed LT:

*‘... evaluative work revealed that the team acknowledged that it was essential to consider the culture of an institution, department and/or discipline as well as the IT infrastructure and support in place’ (2001 p.10)*

### *The EFFECTS project*

The EFFECTS project was a HEFCE-funded exercise in staff development that began with the question of how best to support academics and support staff in HE that wanted to embed communication and information technology (C&IT) into their institutions’ learning and teaching in five HE institutions, three of which were new universities (Harvey and Oliver, 2001 p.7). The original EFFECTS evaluation model was to have feedback from all participants that allowed institutional case studies to be triangulated against snapshots taken across several institutions simultaneously. This proved ineffective because some partners did not contribute much data. Eventually, data provision was linked to continued funding which resulted in enough data for the evaluation to proceed; also annual face-to-face meetings provided an opportunity for contributors to present their experiences directly and with minimal effort (2001 p.5). The report was meant to provide formative evaluation to guide the final year of the project and summative evaluation of the project’s effectiveness in meeting the original aims (2001 p.6). Data was collected for the formative evaluation survey by initial interview with the project teams at each of the institutions followed by questionnaires that were issued to all the stakeholder groups (2001 p.8). Most of the respondents were teachers, the sample contained a high proportion of senior staff and staff support for teaching and learning was more readily available than time or funding. The summative evaluation survey data was drawn from questionnaires involving participants enrolled on EFFECTS programmes at four of the institutions, of which three were new universities. Questionnaires were mailed to 127 contacts and were made available on the WWW. This was followed up with a focus group with representatives from five universities, three of which were new and which investigated personal conceptions relating to embedding C&IT and the EFFECTS programme, followed by exploration of issues raised by previous questionnaire analyses and discussion with EFFECTS project teams. The findings occupy 42 pages, but the summary discussion points out that one of the key outputs was an increase in the confidence of participants in using LT to support learning and teaching (Harvey and Oliver, 2001 p.52).

### *The ASTER project*

Davis et al.’s (2001) final report on the ASTER project has an evaluation section that draws on the unpublished external evaluator’s report (Davis et al., 2001 p.13) which examined three data sources namely written evidence derived from the project that led to the discovery of key themes, interviews with stakeholders and project managers based on those themes and questionnaires returned by academics and support staff who had attended ASTER events. Davis et al. describe how verbal comments were recorded from participants at the pilot workshop and followed by a Web-based questionnaire, the returns from which were disappointing although some participants sent (apparently unsolicited) feedback by email (ibid.). Interestingly, they comment that difficulties in obtaining feedback for the use of such materials are common, and refer to the external evaluator’s view that although the ASTER Web site allowed wide dissemination, the number of hits received by cached web pages does not (for technical reasons) fully reflect use of the site. The external evaluators were reported to ‘advise against further evaluations of this sort’ (ibid.), although the Web site analyses apparently showed external use of the case studies, and it seems that they regarded the project’s case studies as a useful community resource (2001 p.13). The external evaluators also make the point that fixed term funding can exacerbate such inherent difficulties:

*‘...the inflexibility of fixed term funding means that disseminating for embedding, which a development project is, only towards the end of its funding, in a position to undertake, remains unquantifiable’ (2001 p.13)*

Davis et al. argue that the same kind of difficulty is also inherent in measuring such a project's effectiveness and take-up (although it is not clear whether this is meant to be a corollary of fixed term projects), which is, they assert, why they didn't achieve this within the allotted time (2001 p.15). Davis et al. finish by summarising some of the lessons learnt during the ASTER project: Reaching academics within their own institutions was found to be an unexpectedly slow process that was notably more effective when working with specialist learning and teaching centres; LTSN subject centres provide good support for subject specialists; aiming accredited courses at new teachers was effective in reaching new academics; it can be difficult to reach staff who have not yet 'bought into' the benefits of LT; tools for obtaining feedback must be embedded within projects from the beginning; to reach senior management in universities, project outcomes must be presented in a coordinated way, which is difficult but essential if real change in teaching and learning practice is to be brought about- single projects are not enough on their own; publication of findings is a durable, worthy route for dissemination and is both familiar and accessible across the sector; the low profile of research into teaching and learning within disciplines can be a potential barrier; Co-ordinating large numbers of part-time members of a project is difficulties, particularly where there is a need for frequent communication; implementing change takes longer than the lifetime of most projects and fixed time-scale are generally too short to allow implementation and evaluation of change; in addition short-term funding has a negative effect upon all such projects, particularly on staff continuity that results in 'a loss of knowledge and skills' (2001 pp.16-17).

#### *Dempster and Deepwell's metastudy of nationally funded LT projects*

Dempster and Deepwell's metastudy describes the experiences of recent nationally-funded HE LT projects that 'had shown success in embedding new practices into institutional teaching and learning' (Dempster and Deepwell, 2002 p.1), although the criteria for such success are unclear. Six were funded by the Higher Education Funding Council for England (HEFCE) Teaching and Learning Technology Project part 3 (TLTP3), namely TELRI, ASTER, SoURCE (Software Use, Re-Use and Customisation in Education), EFFECTS, TALENT (Teaching And Learning with Network Technologies) and ELEN (Extended Learning Environment Network). ANNIE (Assessing and Networking with National and International Expertise) was funded by the Fund for the Development of Teaching and Learning (FDTL) and three LT projects were funded by ScotCIT: SESDL (Scottish Electronic Staff Development Library), ELICIT (Enabling Large-scale Institutional implementation of C&IT) and NetCulture (Staff Development Network) (Dempster and Deepwell, 2002 p.15). They reiterate criticisms made by Beetham et al. and Davis et al. (Davis et al., 2001; Beetham, 2001) about the disruptive effects on staff of fixed-term funding (2002 pp.2 & 5), and refer to a point made by the Teaching Quality Enhancement Committee (TQEC) that there is a need to address the:

'widespread perception that the arrangements for quality enhancement are complex and fragmented' (TQEC 2003 Cooke Report, quoted by Dempster and Deepwell, 2002 p.2).

In their account of the three national funding bodies (TLTP, FDTL and ScotCIT) Dempster and Deepwell make the point that there are major differences in the extent to which LT was meant to be embedded and outcomes applied outside the context of the funded institution(s), both between the bodies and within different phases of funding within the bodies themselves (2002 pp.2-3). Interestingly, they suggest that in practice, successful embedding a LT may be more a matter of ongoing research and development (R&D) than of simply implementing it (2002 p.3); and quote Taylor's (1998) study of institutional change:

'...the challenge is to move beyond innovation at the level of individual subject or organisational element to change at the institutional level, the reinvention of cultures' (Taylor 1998, quoted by Dempster and Deepwell, 2002 p.3)

The writers refer to the problem of handling the rapid rate of development in LT and identify it as a multidisciplinary field where projects that take place in supportive cultures are most likely to be

successfully embedded. This requires two way communication between academics and LT experts to 'close the loop' when implementing LT and the integrated support of library staff, computing service staff and staff developers, which also resonates with the findings of the national surveys (Bull, 1999; Bull, 2001; Warburton and Conole, 2003). Dempster and Deepwell define such cultures as places where

'...learning technology users do not feel isolated; where the relationship between the centre and the local is strong but flexible; and communications are good.' (2002 p.4).

In a few cases LT projects acted as catalysts for change and transformed institutions at the highest level, but no patterns were identified that may have contributed to this success (2002 p.5). Dempster and Deepwell list the lessons learned from projects that were seen to be successful in embedding LT at the level of the institution: enablers of successful embedding are seen to be academics and support staff 'buying-in' or 'investing in' or 'owning' the innovation; timeliness- the right innovation at the right time in the right institution; collaboration- within the institution, and with other institutions; influential 'champions' of the innovation; integration the innovation into the long-term staff development program; flexibility of interpretation of the project's parameter to accommodate external and internal change (2002 p.6). These observations are also held in common with the findings of the national surveys of CAA (Bull, 1999; Bull and McKenna, 2000; Bull and Hesketh, 2001; Warburton and Conole, 2003). Dempster and Deepwell conclude that the internal dimension of successful change brought about by the projects was largely restricted to ideas and concepts rather than materials and tools, and that the key drivers for successfully embedding LT in HE institutions can be classified as responses to external initiative such as funding programmes, responses to internal strategies and the self-directed responses of the individuals that are driving the project. They suggest that projects which work with a central LT support unit 'with the grain' of the institution are most likely to embed change (2002 p.13).

#### *The CAA Centre*

Two evaluations of the TLTP3 funded CAA Centre were found: Bull's (2001) Final Report as the project's director, and Bull & Conole's (2002) '*Pebbles in the Pond*' external evaluation of the CAA Centre. Bull notes that one of the project's specific objectives was to 'Identify and develop good practice in the embedding of CAA within the curriculum' (Bull, 2001 p.1) and although she does not say in the *Final Report* what constitutes such good practice in CAA, she refers to the *Blueprint* which describes three techniques for embedding CAA (Bull and McKenna, 2001 pp.77-78). Some of the pertinent lessons learnt from the CAA Centre listed by Bull (2001 p.11) are:

- *'The organisational and pedagogical issues and challenges surrounding the take-up of CAA often outweigh the technical limitations of software and hardware.'*
- *Introducing technology into assessment practices requires a lot of perseverance, dedication and resilience. The lack of understanding of the pedagogy of assessment generally, hinders this effort.*
- *Retooling is a challenge which impacts on research and development, requiring a high level of resourcing for academic and support staff in order to maintain pace with technological and software developments.*
- *There is a real need for more funding and research in this area. As on-line learning becomes increasingly common individual institutions do not have the know-how or the resources to progress the pedagogy or the technology to the level at which it will be required. We also perceive that in the future there will be inherent difficulties in teaching and learning on-line and assessing on paper.*
- *Further work should include co-ordination with commercial software suppliers, or better support and long term funding should be available to HE CAA developers, to secure the pedagogical and interoperable foundations and accessibility of software.*

- *More investigation of the cost and time effectiveness of CAA is needed in order to realise the full potential. In addition, integration of CAA with other systems is in early stages and is likely to be where the greatest efficiency gains can be made.*
- *There are political and cultural barriers to implementing CAA which vary between subject areas.*
- *Assessment remains an afterthought in HEFCE funding and research policies for teaching and learning. This verges on naive, given the importance of this activity to the sector, and its potential for legal action.'*

Subject-specific differences in the uptake of CAA are well-known (Bull, 1999; Stephens and Mascia, 1997; Warburton and Conole, 2003) and many of her remaining points about the difficulties inherent in securing permanent change at institutional level persist in CAA and wider LT literature. In particular, her assertion that 'assessment is an afterthought' and that this weakness renders HE institutions vulnerable to appeals and legal action reflects the fears of others (QAA, 1998).

Conole and Bull's paper on the external evaluation of the CAA Centre project included the use of focus groups as a source of data on all aspects of the CAA Centre's remit including the implementation of CAA (2002 p.65). Evidence from these shows that one of the key barriers to the uptake of CAA is the tension caused by the low apparent status of research and development in LT generally and CAA in particular, compared with that of research in the specialist subject areas of academics. Other barriers include subject-specific objections to the use of objective tests and software that is difficult to use (Conole and Bull, 2002 p.70). The provision of high quality technical and pedagogic support is cited as an important enabler of the uptake of CAA, and symmetrically its absence as an important barrier (2002 pp.70-71). They conclude that further work should include researching the cost and time effectiveness of CAA, its integration with VLEs and the subject-specific differences in the application of CAA (2002 p.72).

## Conclusions

This review has reported findings from a survey of the recent literature on CAA that indicate the barriers to embedding CAA in institutions include a lack of commitment at a strategic level, and cultural or pedagogical issues, rather more than any technical shortcomings in the CAA system. Furthermore, computer-assisted assessment in general is still seen as something of an afterthought, and is not yet fully integrated into other learning and teaching practices. Technical and cultural barriers to the adoption of CAA were found to vary between subject boundaries; the enablers to successful embedding of CAA in HE institutions include institutional commitment together with provision of adequate funds, time and personnel, frequent communication between the various stakeholders, interoperability and the existing availability of a VLE.

The literature on institutional embedding of LT (e.g. Dempster and Deepwell, 2002; Taylor, 1998) evidently contributes much to the understanding of how best to embed CAA within HE institutions, and I shall widen the scope of the literature search to include it. Because explicit reference has not been made to all the texts that were found in the search (this article, p.2) this review can not be claimed to be exhaustive. A citation tree must also be constructed in order to identify the seminal articles in this field and to better understand the development of this literature. More research must be carried out to tease apart the different strands that make institutional embedding of CAA most effective and to understand the differences between embedding LT generally, and CAA in particular (Hart, 1998).

**References**

- ALT (2003) The Future of Higher Education: The Association for Learning Technology's response to the White Paper. Found at [www.alt.ac.uk](http://www.alt.ac.uk) on 1-8-03
- Anderson, L., Krathwohl, D. and Bloom, B. (2001) *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Longman, New York.
- Barnett, R. (1997) *Higher education a critical business*, Society for Research into Higher Education & the Open University Press, Buckingham.
- Beetham, H. (2001) Learning technology: institutional audit tools: Notes to JCALT for distribution to all UK HEIs. Found at <http://www.eres.ac.uk/source/docs/pub-ou-35.pdf> on 30/12/2002
- Best, H. (2002) Evaluating the introduction of CBA into the learning, teaching and assessment of the diagnostic radiography course at Sheffield Hallam University. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough.
- Bloom, B., Engelhart, M., Furst, E., Hill, W. and Krathwohl, D. (1956) *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain*, Longman, London.
- Boyle, A., Hutchison, D., O'Hare, D. and Patterson, A. (2002) Item Selection and Application in Higher Education. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough, pp. 269-281.
- Boyle, A. and O'Hare, D. (2003) Assuring quality Computer-Based Assessment development in UK Higher Education. In *7th International CAA Conference*(Ed, Christie, J.) Loughborough University, Loughborough University.
- Brown, G., Bull, J. and Pendlebury, M. (1997) *Assessing Student Learning in Higher Education*, Routledge, London.
- BSI (2002) *Code of practice for the use of information technology (IT) in the delivery of assessments*, BSI, London.
- Bull, J. (1999) Update on the National TLTP3 Project 'The Implementation and Evaluation of Computer-assisted Assessment'. In *3rd International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University.
- Bull, J. (2001) TLTP85 Implementation and Evaluation of Computer-assisted Assessment: Final Report. Found at [http://www.caacentre.ac.uk/dldocs/final\\_report.pdf](http://www.caacentre.ac.uk/dldocs/final_report.pdf) on 16.03.2003
- Bull, J., Conole, G., Davis, H., White, S., Danson, M. and Sclater, N. (2002) Rethinking Assessment Through Learning Technologies. Found at <http://www.ascilite.org.au/conferences/auckland02/proceedings/papers/056.pdf> on 17.May.2003
- Bull, J. and Hesketh, I. (2001) Computer-Assisted Assessment Centre Update. In *5th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University.
- Bull, J. and McKenna, C. (2000) Computer-assisted Assessment Centre (TLTP3) Update. In *4th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University.
- Bull, J. and McKenna, C. (2001) *Blueprint for CAA*, University of Loughborough, Loughborough.
- Conole, G. and Bull, J. (2002) Pebbles in the Pond: Evaluation of the CAA Centre. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University.
- Conole, G. and Oliver, M. (1999) A pedagogical framework for embedding C&IT into the curriculum. In: *ALT-J*, **6**, 4-16.
- Cosemans, D., Van Rentergem, L., Verburgh, A. and Wils, A. (2002) A Campus Wide Setup of Question Mark Perception (V2.5) at the Katholieke Universiteit Leuven (Belgium) - Facing a Large Scale Implementation. In *6th International CAA Conference*(Ed, Danson, M.) University of Loughborough, University of Loughborough.



- Danson, M., Dawson, B. and Baseley, T. (2001) Large Scale Implementation of Question Mark Perception (V2.5) – Experiences at Loughborough University. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University.
- Davies, P. (2001) CAA MUST be more than multiple-choice tests for it to be academically credible? In *5th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough.
- Davies, P. (2002) There's no confidence in multiple-choice testing. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough.
- Davis, M., Oliver, M. and Smith, J. (2001) *The ASTER Project Final Report* London: University of North London, University of London
- Dearing, R. (1997) The National Committee of Enquiry into Higher Education. Found at <http://www.leeds.ac.uk/educol/ncihe/> on 21.04.03
- Dempster, J. and Deepwell, F. (2002) A review of successful project approaches to embedding innovation into institutional teaching and learning practices in higher education. Found at <http://www.warwick.ac.uk/ETS/TELRI/Publications/altbook.pdf> on 26.4.03
- Duke-Williams, E. and King, T. (2001) Using Computer-aided Assessment to test higher level learning outcomes. In *5th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough.
- Farrer, S. (2002) In *Times Higher Education Supplement*.
- Farthing, D. and McPhee, D. (1999) Multiple choice for honours-level students? In *3rd International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough.
- Fink, L. (2003) *Creating Significant Learning Experiences*, Jossey-Bass, San Francisco.
- Gardner-Medwin, A. and Gahan, M. (2003) Formative and Summative Confidence-Based Assessment. In *7th International CAA Conference*(Ed, Christie, J.) Loughborough University, Loughborough.
- Hart (1998) *Doing a literature review releasing the Social Science research imagination*, Sage, London.
- Harvey, J. (Ed.) (1998) *The LTDI Evaluation Cookbook*, Learning Technology Dissemination Initiative, Edinburgh.
- Harvey, J. and Oliver, M. (2001) EFFECTS external evaluation. Found at <http://www.ilrt.org/effects/downloads/effects-evaluation-report.pdf> on 15.May.2003
- Herd, G. and Clark, G. (2003) Computer-aided Assessment- Final Report and Guidelines for Implementation: Implementing CAA in the FE Sector in Scotland. Found at <http://www.onlineassessment.org/Glenrothes%20Report-final.pdf> on 24.8.03
- IMS (2003) (QTI) Enterprise v1.1 Specs, Examples and Schemas. Found at <http://www.imsglobal.org/specificationdownload.cfm> on 2-3-2003
- Kennedy, N. (1998) Experiences of assessing LMU students over the web. Found at <http://www.ulst.ac.uk/cticom/Kennedy.html> on 26.4.03
- Krathwohl, Bloom and Masia (1964) *Taxonomy of educational objectives the classification of educational goals*, Longman, London.
- Laurillard (2002) *Rethinking university teaching a conversational framework for the effective use of learning technologies*, RoutledgeFalmer, London.
- Lay, S. and Sclater, N. (2001) Question and Test Interoperability: An Update on National and International Developments. In *5th International CAA Conference*(Ed, Sclater, N.) Loughborough University, Loughborough.
- Lilley, M. and Barker, T. (2003) An Evaluation of a Computer Adaptive Test in a UK University Context. In *7th International CAA Conference*(Ed, Christie, J.) Loughborough University, Loughborough University.
- McAlpine, M. (2002a) *Design Requirements of a Databank*, CAA Centre, Luton.
- McAlpine, M. (2002b) *Principles of Assessment*, CAA Centre, Luton.

- McAlpine, M. and Hesketh, I. (2003) Multiple Response Questions- Allowing for Chance in Authentic Assessments. In *7th International CAA Conference*(Ed, Christie, J.) Loughborough University, Loughborough.
- McAlpine, M. and Higgison, C. (2001) Online Tutoring e-Book: Chapter 4 New Assessment Methods. Found at <http://otis.scotcit.ac.uk/onlinebook/T4-04.pdf> on 19.8.03
- McCabe, M. and Barrett, D. (2003) CAA Scoring Strategies for Partial Credit and Confidence Levels: CAA- Its a MUGS Game. In *7th International CAA Conference*(Ed, Christie, J.) Loughborough University, Loughborough.
- McKenna, C. and Bull, J. (2000) Quality assurance of Computer-assisted assessment: practical and strategic issues. In: *Quality Assurance in Education*, **8**, 24-31.
- McKenzie, D., Hallam, B., Baggott, G. and Potts, J. (2002) TRIADS experiences and Developments. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough.
- NCFOT (1998) Multiple-Choice Tests. Found at <http://www.fairtest.org/facts/mctfcats.html> on 04.9.03
- O'Leary, R. and Cook, J. (2001) Wading Through Treacle: CAA at the University of Bristol. In *5th International CAA Conference*(Ed, Danson, M.) University of Loughborough, University of Loughborough.
- O'Reilly, M. and Morgan, C. (1999) In *Computer-assisted Assessment in Higher Education*(Eds, Brown, S., Race, P. and Bull, J.) Kogan Page, London, pp. 149-161.
- Parker, A. (2001) TELRI Evaluation. Found at [http://www.warwick.ac.uk/ETS/TELRI/Evaluation/exteval\\_final.doc](http://www.warwick.ac.uk/ETS/TELRI/Evaluation/exteval_final.doc) on 1.5.03
- Paterson, J. S. (2002) What's in a Name? A New Hierarchy for Question Types. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University, pp. 285-292.
- QAA (1998) University of Bath Quality Audit Report, May 1998. Found at <http://www.qaa.ac.uk/revreps/instrev/bath/comms.htm> on 4.9.03
- Raine, J. (1999) Towards the introduction of institution-wide computer assisted assessment: a service department experience. In *3rd International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University.
- Ricketts, C. and Wilks, S. (2001) *is Computer-based Assessment good for students?* Loughborough: Loughborough University
- Sclater, N. and Howie, K. (2003) User requirements of the "ultimate" online assessment engine. In: *Computers in Education*, **40**, 285-306.
- Sclater, N., Low, B. and Barr, N. (2002) Interoperability With CAA: Does It Work In Practice? In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough University, pp. 317-326.
- Shepherd, E. (2001) Assessment Stakes. Found at <http://www.questionmark.com/uk/learningcafe/lowmediumandhighstakesassessments.ppt> on 19.8.03
- Sommerlad, E., Pettigrew, M., Ramsden, C. and Stern, E. (1999) *Synthesis of TLTP annual reports* London: Tavistock Institute
- SQA (2003) *Guidelines for Online Assessment in Further Education* Glasgow: Scottish Qualifications Authority
- Stephens, D. and Mascia, J. (1997) Results of a (1995) Survey into the use of Computer-Assisted Assessment in Institutions of Higher Education in the UK. Found at <http://www.lboro.ac.uk/service/ltf/flicaa/downloads/survey.pdf> on 11-5-2003
- Stevenson, A., Sweeney, P., Greenan, K. and Alexander, S. (2002) Integrating CAA within the University of Ulster. In *6th International CAA Conference*(Ed, Danson, M.) University of Loughborough, University of Loughborough.
- Taylor, P. (1998) Institutional Change in in Uncertain Time: *Lone Ranging* is Not Enough. In: *Studies in Higher Education*, **23**, 269-79.

TOIA (2003) The TOIA Project. Found at [www.toia.ac.uk](http://www.toia.ac.uk) on 31.8.03

Walker, D. and Thompson, J. (2001) A Note on Multiple Choice Exams, with Respect to Students' Risk Preference and Confidence. In: *Assessment & Evaluation in Higher Education*, **26**, pp.261-267.

Warburton, W. and Conole, G. (2003) CAA in UK HEIs: The State of the Art. In *7th International CAA Conference*(Ed, Christie, J.) University of Loughborough, Loughborough, pp. 433-441.

White, S. and Davis, H. (2000) Creating Large Scale Test Banks: a Briefing for Participative Discussion of Issues and Agendas. In *4th International CAA Conference*(Ed, Danson, M.) University of Loughborough, University of Loughborough.

Wood, J. and Burrow, M. (2002) Formative assessment in Engineering using "TRIADS" software. In *6th International CAA Conference*(Ed, Danson, M.) Loughborough University, Loughborough.

Zakrzewski, S. and Steven, C. (2000) A Model for Computer-based Assessment: the catherine wheel principle. In: *Assessment & Evaluation in Higher Education*, **25**, 201-215.