WHITHER E-ASSESSMENT?

Bill Warburton and Gráinne Conole

Whither E-Assessment?

Bill Warburton and Gráinne Conole
University of Southampton
W.I.WARBURTON@soton.ac.uk

Abstract

Assessment is widely regarded as the critical catalyst for student learning (e.g. Brown, Bull and Pendlebury, 1997), and there is considerable pressure on higher education (HE) institutions (HEIs) to measure more formally how learning outcomes (LOs) have been met by students (TQEC Cooke Report quoted by Dearing, 1997; Farrer, 2002; Laurillard, 2002). This has been widely interpreted as a demand for more frequent assessment s a way of better assuring the quality of learning, although few if any additional resources have been available for this: in fact financial resources generally are seen to be static or dwindling. The potential for information and communications technology (ICT) to automate some aspects of learning and teaching (L&T) in HE is widely acknowledged (e.g. Conole, 2004) although promised productivity benefits have been slow to appear (Conole and Dyke, 2004). Computer-aided assessment (CAA) has considerable potential both to ease the assessment load and to provide innovative and powerful assessment modes of assessment in HE (Brown et al., 1997; Bull and McKenna, 2004), and as society shifts inexorably towards technology-based practices there may be *'inherent difficulties in teaching and learning on-line and assessing on paper'* (Bennett, 2002b; Bull, 2001). The article pulls together a number of important strands in the literature and addresses key issues of why CAA isn't more widely used. Links are made between competing pressures on academics and uptake in terms of established diffusion trends.

Introduction

The potential for information and communications technology (ICT) to automate some aspects of learning and teaching (L&T) in HE is widely acknowledged (e.g. Conole, 2004) although promised productivity benefits have been slow to appear (Conole and Dyke, 2004). Computer-aided assessment (CAA) has considerable potential both to ease the assessment load and to provide innovative and powerful assessment modes of assessment in HE (Brown et al., 1997; Bull and McKenna, 2004), and as society shifts inexorably towards technology-based practices there may be *'inherent difficulties in teaching and learning on-line and assessing on paper'* (Bennett, 2002b; Bull, 2001). This begs questions about why CAA isn't more widely used, which metrics are appropriate for ascertaining levels of use and what constitutes good practice when implementing it. The review pulls

together a number of important strands in the literature whilst drawing on diffusion and project risk management scholarship.

Critical factors influencing CAA uptake

After several thousand years, the critical factors influencing the quality and validity of traditional forms of testing such as paper-based or verbal (*viva voce*) assessment might be expected to be well understood by now. However, not all traditional assessments run smoothly (e.g. Goddard, 2002). Many of the barriers and enablers that influence traditional assessment (e.g. security concerns such as candidates seeing questions before the examination) are carried across to CAA. Indeed, the emergence of CAA as a new specialist field of practice appears to have forced the re-examination of dormant issues in traditional practice.

Previous initiatives to identify and classify critical factors in CAA

Several approaches have been taken to identifying and classifying critical factors in technology-based assessment. Derek Stephens and Janine Mascia of Loughborough University conducted the first UK survey of CAA use in 1995, attracting 445 responses from academics. Four years later the CAA Centre's national survey of HE used more detailed versions of Stephens and Mascia's instrument to focus on use and attitudes (Bull, McKenna and Hesketh, 1999). It attracted more than 750 responses from academics, quality assurance (QA) staff and staff developers (McKenna, 2001). Warburton and Conole (2003) ran an adapted, online version of the 1999 survey which received 150 responses, mostly from academic CAA enthusiasts.

Key cultural factors identified were institutional support in terms of training and resourcing, allowing time for academics to develop CAA tests, making CAA a fully integrated part of existing assessment procedures (rather than an afterthought) and subject-related dependencies. Important operational factors reported were student familiarization with the tools, well-planned procedures that address security and reliability issues and cooperative IT departments (Stephens and Mascia, 1997 pp.26-27). Cultural factors identified by academics in 1999 and 2003 were a superset of 1995 findings. 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). 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).

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 higher-order learning outcomes (HLOs)

which resonates with the (2003) findings of Boyle and O'Hare. A credibility gap existed concerning 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. 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 the provision of CAA 'evangelists' and adherence to institutional guidelines was thought less important.

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 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). Most testing was web-based, although a large fraction of respondents delivered CAA using closed networks. Only a third were invigilated, and most of the summative CAA tests restricted the percentage weighting to a third or less, although a small number of CAA tests were worth up to 100% of all marks awarded for a module. Subject-specific differences in the uptake of CAA were obvious (Bull, 1999; Warburton and Conole, 2003). QA staff identified few enabling factors, perhaps indicating a negative perception of CAA (Bull, 1999; Bull and Hesketh, 2001; Bull and McKenna, 2000).

Of the 20 factors identified in 1999 as barriers, 90% were classed as cultural. Of 23 factors identified in 1999 as enablers, about 65% were cultural. Kimberley Hambrick's (2002) Delphi study identified 37 critical factors concerned with large scale applications of formal online assessment in the US K-12 school system, split equally between cultural and operational factors.

Categorising assessment

Much of the pressure on academic and support staff who are running CAA tests derives from the influence that the outcome has on participants' futures (Boyle, Hutchison, O'Hare and Patterson, 2002 p.272; Shepherd, 2001). High stakes assessments are usually seen to be the most sensitive applications of CAA. Shepherd (2001) summarised properties of low, medium and high stakes assessments according to the effect they have on participants' lives (Table 1).

Property of test	Low	Medium	High
Decisions	None	Can be reversed	Difficult to reverse
ID individual	None	Important	Very important
Proctoring	None	Yes	Constant
Options	Study more	Pass, fail, work harder	Pass or fail
Item & test development	Minor	Takes time	Significant
Items created by	Subject expert	Subject expert	Subject expert
Statistics checked	Subject expert	Time to time	Psychometrician

Table 1. Shepherd's (2001) interpretation of summative assessment stakes

Limitations of traditional objective item types

The limitations of existing objective item types, and of their implementation in CAA item design together with the development of new automated item types, are contentious. A major concern related to the nature of objective tests is whether multiple choice questions (MCQs) are really suitable for assessing Higher-order Learning Outcomes (HLOs) in HE students (Davies, 2002) and this is reflected in the opinions of academics and QA staff (Bull, 1999; Warburton and Conole, 2003). MCQs and multiple response questions (MRQ) are still the most frequently used (Boyle et al., 2002; Warburton and Conole, 2003) but there is steady pressure for the use of more sophisticated' question types (Davies, 2001).

Issues of validity and fairness in the use of objective items

Concerns regarding the risk of test-takers achieving passing scores in objective tests by guessing are addressed in two main ways: by discounting a test's guess factor, and 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. In Confidence-based Assessment, marks awarded for a response are predicted on a student's confidence that the correct response has been given (e.g. Davies, 2002).

Concerns about the risk of cheating in summative tests persist (BSI, 2002; Bull and McKenna, 2004; Pain and LeHeron, 2003). Tactics such as collaborative approaches to large banks of items are nothing new in themselves, but may make CAA more worthwhile. Sclater positions item banks as *the* crucial driver of CAA (Herd and Clark, 2003 p.2) whilst McAlpine 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, 2002 p.4).

Boyle et al. (2002) conclude that the present approach of many CAA practitioners to neglect rigorous quality assurance (QA) of items is untenable. They insist that training in item construction and analysis should be obligatory for staff who are involved in developing CAA tests and that items must be peer-reviewed and trialled before use (Boyle et al., 2002 p.77).

Computer-adaptive testing

Concerns about workstation availability and whether objective items explore the limits of a participant's ability are addressed to some degree by computer-adaptive testing (CAT). CAT involves issuing questions of a difficulty level that depends on the test-taker's previous responses. If a question is answered correctly, the estimate of 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. For example, Lilley and Barker (2003) found that 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 (p.177). CAT items are written specifically to test particular levels of ability

and have the potential to deliver results that are more accurate and reliable than simpler tests.

Why isn't uptake higher?

Diffusion theory

The study of ways in which discontinuous innovations such as CAA are taken up is referred to as 'diffusion scholarship' or simply 'diffusion'. The standard reference is probably Emmett Rogers' widely-cited '*Diffusion of Innovations*'. He defines diffusion as the process whereby innovations are taken up within social systems (Rogers, 2003 p.24), which in terms of UK HE could often be on quite a modest scale such as a particular department, school or faculty. Innovations diffuse in five distinct phases, each of which can be identified according to characteristic attributes of users (figure 1).

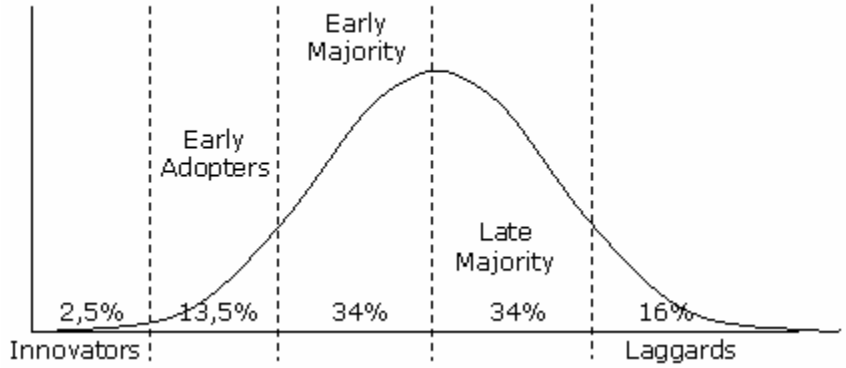

Figure 1. Technology uptake curve (after Rogers, 2003 p.281)

It is argued here that CAA practice in has, in many parts of UK HE, moved beyond 'enthusiasts' in the innovatory and early adoption stages and is at the point of inflection where further uptake depends on opinion formers (Rogers, 2003 pp.26-27). The early majority is characterised by being more risk-averse and pragmatic about innovations than the enthusiasts; although they are not usually technophiles themselves, they tend to respect the judgement of early adopters who are (Rogers, 2003 pp.283-284). In many HE departments, practice is determined by established opinion leaders who see little gain in automating objective tests (Warburton and Conole, 2003).

Geoffrey Moore suggested that a kind of chasm may exist between the early adopters and the early majority. According to this analysis, innovations may fail when the shift to large-scale use reveals underlying weaknesses in the technology which often relate to usability shortcomings (Moore, 1999). The current crop of CAA systems are not generally recognised for their ease of use (Warburton and Conole, 2003).

Risk Management

CAA differs from other learning technologies in that the stakes tend to be higher, particularly where it is used for examinations (QAA, 1998). The clear scoring schemes of objective tests open the results of CAA tests to scrutiny,

rendering deficits in practice visible in ways to which more traditional forms of assessment are not so susceptible. This make thorough risk analysis and management strategies particularly important (Harwood, 2004a; Zakrzewski and Steven, 2000). Operational risks are obvious whereas cultural risks are not. For instance, in Zakrzewski and Steven's sample risk register, only one third of the factors identified could be categorised as cultural (2000, 2003).

Universities host regular teaching quality assurance (TQA) exercises equivalent to OFFSTED inspections of schools. Academics are under pressure to demonstrate that they represent value for money in terms of the value they are able to add to student learning (THES, 2002). At the same time they must raise the profile of their research outputs as assessed by the central funding bodies' research assessment exercise (RAE). Struggling academics may perceive innovation in a focal area of practice such as assessment may entail their taking on more risk than they are prepared to accept (Harwood and Warburton, 2004).

Current activities & debates

Interoperability

Concerns persist concerning whether current CAA systems provide effective levels of interoperability, which is particularly important when preserving the investment made by users of older and newer systems. Another live issue is interoperability between CAA systems and VLEs. 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 candidate for an 'open' CAA *lingua franca* (Lay and Sclater, 2001) but may need considerable development (Sclater and Howie, 2003; Sclater, Low and Barr, 2002). Proprietary standards such as QML exist but usually tie users in to particular products.

Compliance with published standards for CAA practice

The Code of Practice for the Use of Information Technology 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 aims to enhance the status of CAA and encourage its wider use in appropriate applications by demonstrating its fairness, security, authenticity

and validity. 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.77)

Boyle and O'Hare identify three other prescriptions: the American Educational Research Association (AERA) *Standards for Educational and Psychological Testing*, the Association of Test Publishers (ATP) *Guidelines for Computer-based Testing* and the Scottish Qualifications Authority (SQA) *Guidelines for Online Assessment in Further Education* as guides to good CAA practice in as far that they encourage a more all-round treatment of the entire production process including the selection of appropriate question types and analysis of item quality (p.77).

The continuing challenge of scaling up CAA systems

> 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 (Bull, 2001).

An important obstacle to the widespread uptake of CAA is that it is often implemented by individuals or small groups on an *ad hoc* basis with no overarching strategy or IT infrastructure, thus delaying or excluding 'embedding'- uptake on an institutional basis (Boyle and O'Hare, 2003 p.74; McKenna, 2001 p.27). The risks of small-scale development include isolation and underfunding, although benefits include getting things done quickly (Kennedy, 1998). HE institutions that implement CAA centrally encounter risks and benefits on a different scale (Cosemans, Van Rentergem, Verburgh and Wils, 2002; Danson, Dawson and Baseley, 2001; Warburton and Harwood, 2004). Scaling up for full-scale institutional deployment covers every possible use and seems likely to depend more upon the resolution of deployment covers every possible use and seems likely to depend more upon the resolution of cultural than technical issues. Bull pointed out in the (2001) final report of the CAA Centre that

> The organisational and pedagogical issues and challenges surrounding the take-up of CAA often outweigh the technical limitations of software and hardware.

At the level of individual practitioners, perceptions of risk (Harwood, 2004b; Harwood and Warburton, 2004) will continue to play an important part- much of the published research on the large scale implementation of CAA centres on the paramount importance of risk management. Operational obstacles are predictably overcome by incremental advances in technology, but cultural obstacles persist.

Future directions

The increasing sophistication of CAA systems

The shift towards online testing is well documented (e.g. Bennett, 2002a). There have long been demands for CAA systems further to exploit the potential of technology by supporting more flexible questions types that go beyond simple objective item types towards types that are difficult or impossible to rendered on paper (Bennett, 2002b; Bull and Hesketh, 2001; Davies, 2001). These demands have been reflected in the increasing number of question types supported by CAA software: for instance the earlier versions of Questionmark's CAA system supported only eight questions types (Questionmark, 1997) compared with 'more than 20 question types' (Questionmark, 2004) supported currently by it's Perception product. Competitive pressure to produce 'new' question types (sometimes by re-labelling existing types) has led to calls for systematic re-classification of question types (Paterson, 2002).

It seems likely that CAA will continue to require specialist skills: at the level of individual items, known difficulties in question-writing require authors to become experts in the construction of good items (Boyle and O'Hare, 2003). Concerns about cost-benefit may be partially addressed by the integration of other technologies such as multimedia item types that make better use of technology, as indicated by the graphical, animated (FLASH) and open-ended JAVA item types provided by more recent CAA systems (Questionmark, 2004; TOIA, 2004).

Automated essay marking may address concerns about the difficulty of assessing HLOs and is seen by many as the 'Holy Grail' of CAA. It is not (yet) available as part of 'mainstream' systems and requires significant resources in terms of skills and time (Christie, 2003). At the level of system design, pressure for greater interoperability may force the development of standards for test interoperability beyond items, and current de facto standards such as Questionmark Mark-up Language (QML) will be supplemented by the development of 'open' standards such as the IMS Consortium's Question and Test Interoperability (QTI) specification.

Joanna Bull suggested that

> …more investigation of the cost and time effectiveness of CAA is needed in order to realise the full potential…. integration of CAA with other systems is in early stages and is likely to be where the greatest efficiency gains can be made. (Bull, 2001).

Widespread uptake of virtual learning environments (VLEs) and the shift towards managed learning environments (MLEs) may produce systems that are supersets of existing CAA functionality (Brown et al., 1997; Warburton and Conole, 2003) and promise significant usability benefits.

Conclusion

CAA is a vibrant field of research that is maturing in common with e-Learning research generally. The focus of interest is shifting from small-scale, 'cottage industry' applications to larger-scale, more rigorous studies that are more generalisable. It seems likely that the drive towards emergent technologies such as simulations and free-text marking will result in increasingly strong competitive pressures again the more traditional 'standardised testing', purely objective types of CAA system. The shift towards centralised IT systems is likely to be an effective driver for more scalable CAA systems.

References

BENNETT, R. E. (2002a) Inexorable and Inevitable: The Continuing Story of Technology and Assessment. *Journal of Technology, Learning, and Assessment,* 1**,** 1, 3-23.

BENNETT, R. E. (2002b) *Using Electronic Assessment to Measure Student Performance* found at http://www.nasbe.org/Standard/10_Summer2002/bennett.pdf on [17/7/04]

BOYLE, A., HUTCHISON, D., O'HARE, D. & PATTERSON, A. (2002) Item Selection and Application in Higher Education. IN DANSON, M. (Ed.) *6th International CAA Conference.* Loughborough, Loughborough University.

BOYLE, A. & O'HARE, D. (2003) Assuring quality Computer-Based Assessment development in UK Higher Education. IN CHRISTIE, J. (Ed.) *7th International CAA Conference.* Loughborough University, Loughborough University.

BROWN, G., BULL, J. & PENDLEBURY, M. (1997) *Assessing Student Learning in Higher Education,* London, Routledge.

BSI (2002) *Code of practice for the use of information technology (IT) in the delivery of assessments,* London, BSI.

BULL, J. (1999) Update on the National TLTP3 Project 'The Implementation and Evaluation of Computer-assisted Assessment'. IN DANSON, M. (Ed.) *3rd International CAA Conference.* 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 on [16.03.2003]

BULL, J. & HESKETH, I. (2001) Computer-Assisted Assessment Centre Update. IN DANSON, M. (Ed.) *5th International CAA Conference.* Loughborough University, Loughborough University.

BULL, J. & MCKENNA, C. (2000) Computer-assisted Assessment Centre (TLTP3) Update. IN DANSON, M. (Ed.) *4th International CAA Conference.* Loughborough University, Loughborough University.

BULL, J. & MCKENNA, C. (2004) *Blueprint for Computer-assisted Assessment,* London, RoutledgeFalmer.

BULL, J., MCKENNA, C. & HESKETH, I. (1999) Computer-assisted Assessment (CAA) National Survey Questionnaire. CAA Centre, University of Luton.

CHRISTIE, J. (2003) Automated essay marking for content- does it work? IN CHRISTIE, J. (Ed.) *7th International CAA Conference.* Loughborough University, Loughborough.

CONOLE, G. (2004) Keynote: Cultural change- from learners to managers". *FERL Conference 28th January-4th February 2004.* Online, FERL.

CONOLE, G. & DYKE, M. (2004) Where are the affordances of information and communications technology? *ALT-J,* in press.

COSEMANS, D., VAN RENTERGEM, L., VERBURGH, A. & 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 DANSON, M. (Ed.) *6th International CAA Conference.* University of Loughborough, University of Loughborough.

DANSON, M., DAWSON, B. & BASELEY, T. (2001) Large Scale Implementation of Question Mark Perception (V2.5) – Experiences at Loughborough University. IN DANSON, M. (Ed.) *6th International CAA Conference.* Loughborough University, Loughborough University.

DAVIES, P. (2001) CAA MUST be more than multiple-choice tests for it to be academically credible? IN DANSON, M. (Ed.) *5th International CAA Conference.* Loughborough, Loughborough University.

DAVIES, P. (2002) There's no confidence in multiple-choice testing. IN DANSON, M. (Ed.) *6th International CAA Conference.* Loughborough, Loughborough University.

DEARING, R. (1997) *Higher Education in the Learning Society* found at http://www.leeds.ac.uk/educol/ncihe/ on [21.04.03]

FARRER, S. (2002) *End short contract outrage, MPs insist*, The Times Higher Education Supplement [22 November 2002]

GODDARD, A. (2002) *Universities blamed for exam fiasco*, Times Higher Education Supplement [1/11/2002]

HAMBRICK, K. A. (2002) Critical Issues in Online, Large-Scale Assessment: An Exploratory Study to Identify the Issues. Capella University.

HARWOOD, I. (2004a) What happens when Computer-assisted Assessment goes Wrong? *In press*.

HARWOOD, I. (2004b) What happens when Computer-assisted Assessment goes Wrong? *Submitted to British Journal of Educational Technology*.

HARWOOD, I. & WARBURTON, W. (2004) Thinking the Unthinkable: Using Project Risk Management when Introducing Computer-assisted Assessments. IN DANSON, M. (Ed.) *8th International CAA Conference.* Loughborough, Loughborough University.

HERD, G. & 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/cticomp/Kennedy.html on [26.4.03]

LAURILLARD (2002) *Rethinking university teaching a conversational framework for the effective use of learning technologies,* London, RoutledgeFalmer.

LAY, S. & SCLATER, N. (2001) Question and Test Interoperability: An Update on National and International Developments. IN SCLATER, N. (Ed.) *5th International CAA Conference.* Loughborough, Loughborough University.

LILLEY, M. & BARKER, T. (2003) An Evaluation of a Computer Adaptive Test in a UK University Context. IN CHRISTIE, J. (Ed.) *7th International CAA Conference.* Loughborough University, Loughborough University.

MCALPINE, M. (2002) *Design Requirements of a Databank,* Luton, CAA Centre.

MCKENNA, C. (2001) Academic Approaches and Attitudes Towards CAA: A Qualitative Study. IN DANSON, M. (Ed.) *5th International CAA Conference.* Loughborough, Loughborough University.

MOORE, G. A. (1999) *Crossing the chasm,* Oxford, Eng., Capstone Publishing.

PAIN, D. & LEHERON, K. (2003) WebCT and Online Assessment: The best thing since SOAP? *Educational Technology & Society,* 6**,** 2, 62-71.

PATERSON, J. S. (2002) What's in a Name? A New Hierarchy for Question Types. IN DANSON, M. (Ed.) *6th International CAA Conference.* Loughborough University, Loughborough University.

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]

QUESTIONMARK (1997) *Overview of Question Mark for Windows* found at http://www.questionmark.com/us/qmwindows/manuals/qmforwindows/section 6/6.html#6.1%20Overview on [29-8-2004]

QUESTIONMARK (2004) *Perception: Windows Based Authoring* found at http://www.questionmark.com/uk/perception/authoring_windows.htm on [29-8-2004]

ROGERS, E. M. (2003) *Diffusion of Innovations,* New York, Free Press.

SCLATER, N. & HOWIE, K. (2003) User requirements of the ‘‘ultimate’’ online assessment engine. *Computers in Education,* 40**,** 3, 285–306.

SCLATER, N., LOW, B. & BARR, N. (2002) Interoperability With CAA: Does It Work In Practice? IN DANSON, M. (Ed.) *6th International CAA Conference.* Loughborough University, Loughborough University.

SHEPHERD, E. (2001) *Assessment Stakes* found at http://www.questionmark.com/uk/learningcafe/lowmediumandhighstakesassessments.ppt on [19.8.03]

STEPHENS, D. & 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/ltd/flicaa/downloads/survey.pdf on [11-5-2003]

THES (2002) *'Cheque-book entry' rejected*, THES (Times Higher Education Supplement) [06 December 2002]

TOIA (2004) *The TOIA Project* found at www.toia.ac.uk on [31.8.04]

WARBURTON, W. & CONOLE, G. (2003) CAA in UK HEIs: The State of the Art. IN CHRISTIE, J. (Ed.) *7th International CAA Conference.* Loughborough, University of Loughborough.

WARBURTON, W. & HARWOOD, I. (2004) Implementing Perception across a large University: getting it right. *Perception European Users Conference.* Edinburgh, Question mark Computing.

ZAKRZEWSKI, S. & STEVEN, C. (2000) A Model for Computer-based Assessment: the catherine wheel principle. *Assessment & Evaluation in Higher Education,* 25**,** 2, 201-215.

ZAKRZEWSKI, S. & STEVEN, C. (2003) Computer-based Assessment: quality assurance issues, the hub of the wheel. *Assessment & Evaluation in Higher Education,* 28**,** 6, 609-623.