

Implementation of computer assisted assessment: lessons from the literature

Gavin Sim*, Phil Holifield & Martin Brown

University of Central Lancashire, UK

This paper draws attention to literature surrounding the subject of computer-assisted assessment (CAA). A brief overview of traditional methods of assessment is presented, highlighting areas of concern in existing techniques. CAA is then defined, and instances of its introduction in various educational spheres are identified, with the main focus of the paper concerning the implementation of CAA. Through referenced articles, evidence is offered to inform practitioners, and direct further research into CAA from a technological and pedagogical perspective. This includes issues relating to interoperability of questions, security, test construction and testing higher cognitive skills. The paper concludes by suggesting that an institutional strategy for CAA coupled with staff development in test construction for a CAA environment can increase the chances of successful implementation.

Introduction

This paper presents evidence that the more traditional methods of assessment within universities have their limitations. As a result of these limitations and also the continued increase in the use of technology to deliver curriculum, the gap between assessment methods and learning is widening.

Students entering higher education directly from schools and colleges are likely to have been exposed to Information Technology as part of the UK National Curriculum. Pilot studies conducted within schools for the delivery of summative assessment via the web (Ashton *et al.*, 2003; Nugent, 2003) and for basic key skills tests in both Learn Direct and army centres (Sealey *et al.*, 2003) indicate that CAA can successfully assess students and provide timely feedback regarding class and individual progress. There is also empirical evidence to suggest students find CAA an acceptable assessment technique (Sambell *et al.*, 1999; Croft *et al.*, 2001; Ricketts & Wilks, 2002a). Therefore it could be argued that for many students CAA may become a

*Corresponding author: Department of Computing, Computing and Technology Building, University of Central Lancashire, Preston PR1 2HE, UK. Email: grsim@uclan.ac.uk

more widely used method of assessment in schools, Further education and Universities. Many universities are now using technology in their assessment strategies (Stephens & Mascia, 1997) and by examining the literature, lessons can be learned to facilitate the successful implementation of computer-assisted assessment.

Methodology

A comprehensive literature review was conducted primarily using online resources, although library resources were also used. The searching centred on the databases Ingenta, AACE, Science Direct and Conference Proceedings such as the International Computer Assisted Assessment Conference. Key word searching was problematic and time consuming, for example conducting a search using 'computer assessment' would produce a divergent array of articles in excess of one thousand. Other terminology used in the search included 'computer based testing, computer-based assessment, computer aided assessment and e-assessment'. Browsing through the contents of entire journals such as 'Assessment and Evaluation in Higher education' was also adopted.

Assessment in general

Academic assessment can be administered through various techniques. Fifty varied techniques have been identified and used within higher education for assessment purposes (Knight, 2001); the most commonly used are exams and essays (Graham, 2004). However this does not include all the methods now available within CAA packages for example incorporate questions that make use of multimedia. New assessment techniques will continue to emerge as technology and teaching methods change and develop, therefore continuing research will be required to determine the effectiveness and appropriateness of these methods.

Each form of assessment presents its own difficulties, whether computer based or traditional. Essays present the problem of double marking, in one study both markers agreed only 52% of the time (Powers *et al.*, 2002). Additionally there are the problems with cheating as Internet sites offer custom-written and off the shelf essays (Crisp, 2002). It has been suggested that exams tend to encourage surface learning (Race, 1995) and may cause increased anxiety resulting in significantly lower scores (Cassady & Johnson, 2002). The multiple choice question (MCQ) styles are used in both offline and CAA exams and raise a number of concerns, for example, grade deflation by not enabling partial credit (Baranchik & Cherkas, 2000), poorly designed questions (Paxton, 2000; Jafarpur, 2003) and guessing (Burton, 2001). However the advantages of using computers to deliver MCQ for lecturers include automated marking (Pollock *et al.*, 2000) and for formative purposes the students have the opportunity to study at their own pace, repeat questions and receive instant feedback (Loewenberger & Bull, 2003). It is the potential advantages of CAA that has driven research into ways to overcome the difficulties.

Ultimately in an academic environment the marks from summative assessment are accumulated to award an overall grade and there are concerns over comparability across

subject domains. It has been suggested that the scientific subjects produce more First Class Degrees than the humanities because of the nature of the marking criteria in using the full range of marks and subjectivity is eliminated from the equation where there is a predefined correct answer (Yorke *et al.*, 2002; Horney, 2003). These findings would appear to be further corroborated by the Higher education Statistics Agency (HESA) figures. Of the students graduating from UK universities in 2001/02 in Mathematical Science 25.5% passed with a First Class Degree, compared to 10.4% in Humanities (HESA, 2002) and this trend was also evident in other years for example, 1994/95 (HESA, 1995). CAA, like mathematics and some science subjects, also tends to use the full range of marks therefore the trend towards a high proportion of First Class Degrees may occur in other subject domains adopting this technique in the future.

There is pressure on lecturers not to fail students, and one study found that in professional subjects there is a tendency to leave the award of a fail to the next assessor (Hawe, 2003). Lecturers are confronted with emotional and ethical dilemmas when a close working relationship is formed, increasing their reluctance to award a fail (Sabar, 2002). The emotional and subjectivity issues that are evident in human centred marking may be removed via automatic marking offered by CAA software.

It is important to recognize that some of these issues discussed are still prevalent in CAA along with new challenges. Adopting a diverse assessment strategy may lead to a fairer assessment of the student (Race, 1995).

Computer-assisted assessment defined

From the literature there is a lack of universal consent regarding the terminology and its definition, however, Bull and McKenna (2001) argue that computer-assisted assessment is the common term for the use of computers in the assessment of students and the other terminology tend to focus on the activities. Therefore the definition of CAA used in this review will be that: CAA encompasses the use of computers to deliver, mark or analyse assignments or exams.

Variations in CAA

Within higher education institutions the application of CAA has occurred in a number of varied ways, these include, adaptive testing (Latu & Chapman, 2002; Mills *et al.*, 2002), analysis of the content of discussion boards (Macdonald & Twining, 2002; Wiltfelt *et al.*, 2002), automated essay marking (Christie, 1999; Burstein *et al.*, 2001), delivery of exam papers (Sim *et al.*, 2003) and objective testing (Walker & Thompson, 2001; Pain & Le Heron, 2003). These methods vary considerably however the focus of this review of research will centre on the issues relating to implementing objective tests via CAA.

Testing cognitive skills with CAA

There is concern in the literature relating to CAA and its ability to test higher cognitive skills across subject domains (Daly & Waldron, 2002; Paterson, 2002). The higher

cognitive skills are often associated with 'Analysis, Synthesis, and Evaluation' as defined in Bloom's Taxonomy (Bloom, 1956). However, a revised taxonomy takes into consideration the 'Knowledge Dimension' (Anderson & Krathwohl, 2001) and this has also been used in CAA research for classification of questions (King & Duke-Williams, 2002; Mayer, 2002).

Paterson (2002) indicated that it is not feasible to test the higher-level cognitive skills using CAA within mathematics. Bloom states that in the majority of instances Synthesis and Evaluation promote divergent thinking and answers cannot be determined in advance (Bloom *et al.*, 1971). Heinrich and Wang (2003) argue that objective testing is still not sophisticated enough to examine complex content and thinking patterns. However, other research in linguistics and computer programming concluded that the higher-level skills can be assessed via CAA through innovative approaches (Cox & Clark, 1998; Reid, 2002). In the study by Reid (2002) a new language was devised and students were required to apply linguistic techniques in order to answer MCQ. It has been suggested that CAA tests of higher-level skills are more complex and costly to produce (Dowsing, 1998) and this may be because more innovative approaches are needed.

Question styles

Objective testing has been used within assessment for over forty years (Wood, 1960) and computer programs delivering MCQ date back to the 1970s (Morgan, 1979). More sophisticated question styles have emerged enabling more diverse assessment methods. The question styles delivered by the TRIADS software developed at Derby University are evidence of this evolution, offering 17 question styles in 1999 (Mackenzie, 1999) and 39 in 2003 (CIAD, 2003). However, staff at the University of Liverpool using TRIADS found that this presented an additional problem, as they were unfamiliar with the new question styles and lacked confidence in writing suitable questions (McLaughlin *et al.*, 2004). Staff development in writing suitable questions and guidelines can be used to overcome these problems. For example, generic guidelines developed by Haladyna (1996), Herd and Clark (2002) present examples of the various questions styles used in further education whilst examples used within higher education can be found at <http://www.caacentre.ac.uk>.

Although there are a large number of possible formats for CAA questions, it is possible to classify them into four distinct groups based on the human interaction technique required (CIAD, 2003). These groups are defined as point and click, move object, text entry and draw object.

Point and click

Point and click questions include Multiple Choice (MCQ) and Multiple Response (MRQ) items, which have both been used within assessment practise for a considerable time and as a result are often transformed into CAA (Ricketts & Wilks, 2002b). Ebel (1972) suggests that any understanding or ability that can be tested by means of

any other technique, for instance essays, can also be tested by MCQ. More complex MCQ questions can be devised through assertion reasoning resulting in the testing of higher cognitive skills (Bull & McKenna, 2001). Both MCQ and MRQ have inherent problems, such as reliance on true and false style questions which students might perceive to be unfair (Wood, 1960). Davies also argues that the quality of MCQ is dependent on the quality of the distracter and not the question (Davies, 2002).

Move object

Move object style questions focus on the movement of objects to predetermined positions on the screen. They are a variation of the MCQ format and are good for assessing students understanding of relationships (Bull & McKenna, 2001). For example in computing they could be used for the labelling of entity relationship diagrams or in linguistics students could be presented with a poem and move the highlighted words to the appropriate word class. One problem is that when the number of moveable objects is equal to the number of targets, if a student knows all but one answer they will automatically get full marks (Wood, 1960).

Text entry

Text entry questions consist of input of short predefined answers, such as factual knowledge or syntax in computer programming. An advantage of this format is that students must supply the correct answer removing the possibility of guessing (Bull & McKenna, 2001) and this style has been found to be the most demanding format for students (Reid, 2002). There are problems associated with text entry within some subject domains such as mathematics, as mathematical expressions cannot easily be included in most commercial software (Croft *et al.*, 2001; Paterson, 2002). Another problem associated with this question style is that the answer may be marked incorrect due to spelling mistakes and the time saving element may be reduced if lecturers need to manually check for spelling errors.

Draw object

This is associated with drawing simple objects or lines. For example, students may be required to plot graphs which can be automatically marked. This style of question is a high discriminator between strong and weak candidates (Mackenzie, 1999). There is little evidence in the literature concerning the effectiveness of this format, but this might be due to the fact that commercial software such as Questionmark and I-Assess do not have this style in their templates.

Interoperability and question banks

Question banks which are authored and peer reviewed by academics are emerging, such as the Electrical and Electronic Engineering Assessment Network who

developed a database of questions in electrical and electronic engineering (Bull *et al.*, 2002). One such bank will typically require 5000 questions making it unfeasible for a single institution to develop (Maughan *et al.*, 2001). Constructing high quality questions is difficult, time consuming and expensive (Sclater *et al.*, 2003) and issues arise in the interoperability of questions between CAA Software (Lay & Sclater, 2001). There are several international standards established to enable interoperability of questions between software applications (Herd & Clark, 2003). These specifications are based on metadata structure for questions and their grouping together. Unless these interoperability standards are developed and utilized question banks will have a limited life, as they cannot be used on a variety of delivery platforms (White & Davis, 2000). Systems are emerging that are IMS-QTI compliant (Instructional Management Systems – Question and Test Interoperability Specification) to facilitate the exchange of questions (Daly, 2002; Bacon, 2003). The Centre for Educational Technology Interoperability Standards (www.cetis.ac.uk) offers comprehensive resources and information on the issues concerning interoperability which may help direct further research.

Guessing

A number of the question styles associated with CAA can lead to artificially high marks through guessing (Bush, 1999), which has implications for setting the pass mark of the test. For example, setting a pass mark of 40% based on assessment of true/false answers would be inappropriate, as guessing alone would give an average of 50% (Harper, 2002). The problems of guessing may be addressed through various marking schemes, such as post test correction (Bull & McKenna, 2001), negative marking (Bush, 1999), increasing the number of questions or combining the results from several tests (Burton & Miller, 1999) or increasing the number of distracters and the pass mark (Mackenzie & O'Hare, 2002). It has been suggested that negative marking is not generally implemented in the UK (McAlpine, 2002) and that post test correction is only suitable with a single question style because the formulae would vary depending on the number of distracters (Harper, 2003).

Statistical analysis has resulted in various methods being developed to assist in test construction in order to reduce the effects of guessing. An empirical marking simulator to assist in scoring and test construction based on a base level guess factor has been developed (Mackenzie & O'Hare, 2002), this program examines the mark distribution and measurement scale for a set of random answers, enabling tutors to establish the effects of guessing on their assessment. Also statistics to award a score for partial credit through a formula based on a mean uneducated guessers score has been investigated (McCabe & Barrett, 2003). This allows MCQ to be unconstrained, similar to MRQ styles, enabling students to provide more than one answer and their score is weighted depending on the number of choices. For example, an MCQ with one correct answer, four possible options and a score of 3, if a student includes the correct answer by selecting 2 options they would only score 2 ($2=3-1$). Davies used a combination of predetermining the students' confidence in answering the question

prior to seeing the distracters and negative marking, resulting in students perceiving this to be a fairer test of their abilities (Davies, 2002).

There is lack of evidence that any one specific technique generates more accurate results than any other. It could be argued that these techniques are unnecessary if the tests are well constructed (Bull & McKenna, 2001).

Accessibility

UK institutions now have to comply with the Special Educational Needs and Disability Act when preparing both teaching and assessment material (SENDA, 2001). The number of students in UK higher education registering a disability in 2000 was 22,290 and this has implications for CAA (Phipps & McCarthy, 2001). For example, a student with dyslexia may exert more cognitive resources in interpreting the question, therefore, ensuring the language is appropriate is a necessity (Wiles & Ball, 2003). In addition extra time may be required to complete the test which may necessitate the publishing of two different assessments, one with a longer duration. Feedback from one dyslexic student regarding CAA indicated that they thought it provided a more level playing field in which they can demonstrate their knowledge (Jefferies *et al.*, 2000). Students with visual or physical impairment may struggle to answer move object and draw object style questions without the aid of assistive technology, they may need specially adapted input software and hardware such as, touch screens, eyegaze systems or speech browsers.

There are guidelines for general teaching, however there is little evidence that guidelines for inclusive and accessible design in CAA are emerging (Wiles, 2002). For example, when multimedia elements, such as video are used within the assessment, it may necessitate the provision of an alternative paper-based version for students with sensory impairment. The introduction of an alternative, in this instance paper, poses the problem of ensuring comparability (Bennett *et al.*, 1999). When identical tests are presented on a computer and paper they are not comparable (Clariana & Wallace, 2002) because there are numerous variables that impact on student's performance when questions are presented on a computer. These variables include the monitor (Schenkman *et al.*, 1999), the way text is displayed on screen (Dyson and Kipping, 1997), reading from a monitor is slower than paper (Mayes *et al.*, 2001) and the problems of obtaining a feel for the exam when only a single question is presented (Liu *et al.*, 2001). The Web Accessibility Initiative (<http://www.w3c.org/WAI/>) has produced useful guidelines for promoting online accessibility which may be applicable to CAA but this initiative does not address the issue of comparability between questions.

Institutional strategies for the adoption of CAA

The greatest barrier to the adoption of CAA by academics is lack of time, to both develop questions and learn the software (Warburton & Conole, 2003). This may have contributed to the fact that the adoption of CAA has usually resulted from the

impetus of enthusiastic individuals rather than strategic decisions (O'Leary & Cook, 2001; Daly & Waldron, 2002). The perceived benefits of CAA of freeing lecturers' time can be illusive if no institutional strategy or support is offered (Stephens, 1994), successful implementation may be left to chance (Stephens *et al.*, 1998) and CAA may be developed in an anarchic fashion (McKenna & Bull, 2000). Research conducted at the University of Portsmouth indicate that there is no time saving benefit for courses with less than twenty students (Callear & King, 1997). In order to utilize the features within software packages staff training and development is necessary (Boyle & O'Hare, 2003) and this may not be feasible without institutional support.

Institutions adopting CAA are faced with the difficulty of evaluating and deciding upon the most appropriate CAA software. Without an institutional strategy, individual departments may adopt their own systems (O'Leary & Cook, 2001). This results in students having to cope with a number of different user interfaces and CAA formats, increased licence costs and problems offering administrative and technical support. Even if an institution has a clear strategy there are also problems in determining the selection criteria for software used to deliver assessment and there is a lack of analysis within the literature (Valenti *et al.*, 2002). Sclater and Howie (2003) contributed to this literature by defining the ultimate online assessment engine. This was achieved through a process of examining the user requirements of the system, establishing the stakeholders and their functional requirements. This research may aid institutions identify their needs and establish an appropriate evaluation methodology.

The following guidelines for an institutional strategy have been formulated by Loughborough University and the University of Luton: establish a coordinated CAA management policy for CAA unit(s) and each discipline on campus; establish a CAA unit; establish CAA discipline groups/committees; provide funding; organize staff development programmes; establish evaluation procedures; identify technical issues; establish operational and administrative procedures (Stephens *et al.*, 1998). BS7988 is a new British Standard Code of practice that has been introduced governing the use of information technology in the delivery of assessments (BS7988, 2002). The guidelines have various implications for the delivery of assessments, for example, it is recommended that students take a break after 1.5 hours which has an impact on the invigilation process. If this recommendation is followed, procedures need to be established to prevent collusion between students during the break or the tests need to be split into two separate sections. One of the difficulties for many institutions using CAA arises through the lack of resources to accommodate large cohorts of students sitting the exam simultaneously (Mackenzie *et al.*, 2004). This problem can be alleviated through institutional support and therefore, to fully utilize the benefits of CAA an institutional strategy would appear necessary to increase the chance of successful implementation. These benefits are evident within a number of institutions with strategies, such as, Ulster (Stevenson *et al.*, 2002), Derby (Mackenzie *et al.*, 2002), Coventry (Lloyd *et al.*, 1996) and Loughborough (Croft *et al.*, 2001).

Security

The move from traditional teaching environments and examination settings presents additional issues relating to security. Frohlich (2000) states that in traditional environments it is possible to ensure the security of the exam papers and scripts, this includes the transportation to and from the exam venue. However, even under this system breaches in security do occur, for example AQA had to replace 500,000 English and English Literature exam papers after a box had been tampered with (Curtis, 2003).

Tannenbaum (1999) defines security in computer systems as consisting of procedures to ensure that individuals cannot access material for which they do not have authorisation. This is essential within a CAA environment as questions and student details are stored in a database and usually the test data is sent over a local network or the Internet. Before computers were connected to the Internet it was relatively easy to have effective security measures (Mason, 2003), but transmission of sensitive data over an insecure network requires additional security measure to be implemented.

Encryption techniques can be used to ensure the security of the questions and answers when transmitting data over the Internet (Sim *et al.*, 2003). To increase security, examinations can be loaded on to the server at the last minute (Whittington, 1999). If email is used to submit results there is a potential risk due to the lack of authentication (Hatton *et al.*, 2002). Four security requirements have been identified by Luck and Joy, these being: all submissions must be logged, it must be verified that a stored document used for the assessment is the same as the one used by the student, a feedback mechanism must inform students that their submission has been received and the identity of the student must be established (Luck & Joy, 1999).

With the majority of CAA software students and administrators are required to have passwords which is often the weakest link in terms of protection (Hindle, 2003). Although an unlikely event, students could get access to the administrator password and change their results or gain access to the questions. Other concerns include authentication and invigilation of the students, which can be particularly problematic in remote locations (Thomas *et al.*, 2002). At present students enrolled on distance learning courses overseas need to sit exams in a specific location such as, the British Council Offices to enable authentication and invigilation. Research is being conducted to overcome these problems but unless solutions are found, geographical barriers will remain as students need access to the test centres.

During the test computers need to be locked down, removing the possibility of accessing other content and secure browsers have been developed to enable this such as, Questionmark Secure (Kleeman & Osborne, 2002). There are operational risks associated with CAA that have security implications such as the server crashing and these risks need to be identified and procedures established to minimize them (Zakrzewski & Steven, 2003).

There are software standards for security for example, the British Standards on Information Security Management BS7799, which has also been adopted as an

International Standard IS17799. In addition, when data from the test has been collected institutions within the UK should abide by the Data Protection Act 1998 (Mason, 2003). If security measures are in place there is no evidence to suggest that the integrity of the examination is more compromised by delivery over the Internet than by paper.

Conclusion

The implementation of CAA from a technical and pedagogical perspective is a complex process. The first, and perhaps the most important, lesson that can be learned is that an institutional strategy would seem to greatly increase the chances of success. There are recommendations that have been made to assist policy makers formulate an effective strategy. Without institutional support implementing security procedures may be more problematic, such as locking down PCs. However, authentication and invigilation in remote locations is still an issue that has yet to be fully resolved.

The other important lesson that can be learned is in relation to staff development and training in test construction within a CAA environment. Focused staff development may help alleviate a number of issues, such as guessing, testing various cognitive skills, using appropriate question styles and accessibility. The emergence of question banks may also address these issues depending on their level of interoperability. Another issue is that whilst there are guidelines relating to accessible online content there are still no formal guidelines relating to CAA.

The reliance on a single method of assessment is problematic and a diverse assessment strategy is usually necessary. Within an environment of increasing student numbers and a reduction of staff to student ratio, CAA would appear to be a partial solution. This study has highlighted the issues surrounding the implementation of CAA to both inform and direct further research in the field.

References

- Anderson, L. W. & Krathwohl, D. R. (2001) *A taxonomy for learning, teaching, and assessing. A revision of blooms taxonomy of educational objectives* (New York, Longman).
- Ashton, H. S., Schofield, D. K. & Woodger, S. C. (2003) Pilot summative web assessment in secondary education, *Proceedings of the 7th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 19–29.
- Bacon, R. A. (2003) Assessing the use of a new QTI assessment tool within Physics, *Proceedings of the 7th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 33–44.
- Baranchik, A. & Cherkas, B. (2000) Correcting grade deflation caused by multiple-choice scoring, *International Journal of Mathematical Education in Science and Technology*, 31(3), 371–380.
- Bennett, R. E., Goodman, M., Hessinger, J., Kahn, H., Liggett, J., Marshall, G. & Zack, J. (1999) Using multimedia in large-scale computer-based testing programs, *Computers in Human Behaviour*, 15(3), 283–294.
- Bloom, B. S. (1956) *Taxonomy of educational objectives: the classification of educational goals. Handbook 1. Cognitive domain* (New York, Longman).

- Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971) *Handbook on formative and summative evaluation of student learning* (New York, McGraw-Hill Books).
- Boyle, A. & O'Hare, D. (2003) Finding appropriate methods to assure quality computer-based assessment development in UK higher education, *Proceedings of the 7th International Computer Assisted Assessment Conference*, (Loughborough, Loughborough University), 67–82.
- BS7988 (2002) *Code of practice for the use of information technology (IT) in the delivery of assessment*
- Bull, J., Conole, G., Davis, H. C., White, S., Danson, M. & Sclater, N. (2002) Rethinking assessment through learning technologies, *Proceedings of the ASCILITE 2002* (Auckland, UNITEC), 1–12.
- Bull, J. & McKenna, C. (2001) *Blueprint for computer-assisted assessment* (Loughborough, Loughborough University).
- Burstein, J., Leacock, C. & Swartz, R. (2001) Automated evaluation of essays and short answers, *Proceedings of the 5th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Burton, R. F. (2001) Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers, *Assessment and Evaluation in Higher Education*, 26(1), 41–50.
- Burton, R. F. & Miller, D. J. (1999) Statistical modelling of multiple-choice and true/false tests: ways of considering, and of reducing, the uncertainties attributed to guessing, *Assessment and Evaluation in Higher Education*, 24(4), 399–411.
- Bush, M. (1999) Alternative marking schemes for on-line multiple choice tests, *Proceedings of the 7th Annual Conference on the Teaching of Computing* (Belfast, Elsevier).
- Callar, D. & King, T. (1997) Using computer-based tests, *ALT-J*, 5(1), 27–32.
- Cassady, J. C. & Johnson, R. E. (2002) Cognitive test anxiety and academic performance, *Contemporary Educational Psychology*, 27(2), 270–295.
- Christie, J. R. (1999) Automated essay marking for both content and style, *Proceedings of the 3rd Annual Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- CIAD (2003) Summary of question styles. Available online at: <http://www.derby.ac.uk/ciad/ciastyles.html> (accessed 30 June 2003).
- Clariana, R. & Wallace, P. (2002) Paper-based versus computer-based assessment: key factors associated with test mode effect, *British Journal of Educational Technology*, 33(5), 593–602.
- Cox, K. & Clark, D. (1998) The use of formative quizzes for deep learning, *Computers and Education*, 30(3), 157–167.
- Crisp, B. R. (2002) Assessment methods in social work education: a review of the literature, *Social Work Education*, 21(2), 259–269.
- Croft, A. C., Danson, M., Dawson, B. R. & Ward, J. P. (2001) Experience of using computer assisted assessment in engineering mathematics, *Computers and Education*, 37(1), 53–66.
- Curtis, P. (2003) *Missing paper sparks exam reprint* (London, Guardian).
- Daly, C. & Waldron, J. (2002) Introductory programming, problem solving and computer assisted assessment, *Proceedings of the 6th International Computer Assisted Assessment Conference*, (Loughborough, Loughborough University), 95–106.
- Daly, J. (2002) An XML question bank using Microsoft Office, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 107–118.
- Davies, P. (2002) There's no confidence in multiple-choice testing, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 119–132.
- Dowsing, R. D. (1998) Flexibility and the technology of computer aided assessment, *Proceedings of the ASCILITE 1998* (Wollongong, University of Wollongong), 163–171.
- Dyson, M. C. & Kipping, G. J. (1997) The legibility of screen formats: are three columns better than one? *Computers and Graphics*, 21(6), 703–712.
- Ebel, R. L. (1972) *Essentials of educational measurement* (Englewood Cliffs, Prentice-Hall).

- Frohlich, R. (2000) Keeping the wolves from the door, wolves in sheep clothing, that is, *Proceedings of the 4th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Graham, D. (2004) A survey of assessment methods employed in UK higher education programmes for HCI courses, *Proceedings of the 7th HCI Educators Workshop* (Preston, LTSN), 66–69.
- Haladyna, T. M. (1996) *Writing test items to evaluate higher order thinking* (Needham Heights, Allyn & Bacon).
- Harper, R. (2002) Allowing for guessing and for expectations from the learning outcomes in computer-based assessments, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 139–150.
- Harper, R. (2003) Correcting computer-based assessment for guessing, *Journal of Computer Assisted Learning*, 19(1), 2–8.
- Hatton, S., Boyle, A., Byrne, S. & Wooff, C. (2002) The use of PGP to provide secure email delivery of CAA results, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 149–160.
- Hawe, E. (2003) It's pretty difficult to fail; the reluctance of lecturers to award a fail grade, *Assessment and Evaluation in Higher Education*, 28(4), 371–382.
- Heinrich, E. & Wang, Y. (2003) Online marking of essay-type assignments, *Proceedings of the World Conference on Educational Multimedia Hypermedia and Telecommunications* (Hawaii, AACE), 768–772.
- Herd, G. & Clark, G. (2002) *Computer assisted assessment implementing CAA in FE sector in Scotland: question types* (Glenrothes, Glenrothes College).
- Herd, G. & Clark, G. (2003) *CAA implementation in the FE Sector in Scotland* (Glenrothes, Glenrothes College).
- HESA (1995) *Students in higher education institutions 1994/95* (London, HMSO).
- HESA (2002) *Students in higher education institutions 2001/02* (London, HMSO).
- Hindle, S. (2003) Careless about privacy, *Computers and Security*, 22(4), 284–288.
- Horney, W. (2003) Assessing using grade-related criteria: a single currency for universities? *Assessment and Evaluation in Higher Education*, 28(4), 435–454.
- Jafarpur, A. (2003) Is the test constructor a facet? *Language Testing*, 20(1), 57–87.
- Jefferies, P., Constable, I., Kiely, B., Richardson, D. & Abraham, A. (2000) Computer aided assessment using WebCT, *Proceedings of the 4th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- King, T. & Duke-Williams, E. (2002) Using computer aided assessment to test higher level learning outcomes, *Proceedings of the 5th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Kleeman, J. & Osborne, C. (2002) A practical look at delivering assessment to BS7988 recommendations, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 163–170.
- Knight, P. (2001) *A briefing on key concepts formative and summative, criterion and norm-referenced assessment* (York, LTSN Generic Centre).
- Latu, E. & Chapman, E. (2002) Computerised adaptive testing, *British Journal of Educational Technology*, 33(5), 619–622.
- Lay, S. & Sclater, N. (2001) Question and test interoperability: an update on national and international developments, *Proceedings of the 5th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Liu, M., Papathanasiou, E. & Hao, Y. (2001) Exploring the use of multimedia examination formats in undergraduate teaching: results from the fielding testing, *Computers in Human Behaviour*, 17(3), 225–248.
- Lloyd, D., Martin, J. G. & McCaffery, K. (1996) The introduction of computer based testing on an engineering technology course, *Assessment and Evaluation in Higher Education*, 21(1), 83–90.

- Loewenberger, P. & Bull, J. (2003) Cost-effectiveness analysis of computer-based assessment, *ALT-J*, 11(2), 23–45.
- Luck, M. and Joy, M. (1999) A secure on-line submission system, *Software—Practice and Experience*, 29(8), 721–740.
- Macdonald, J. & Twining, P. (2002) Assessing activity-based learning for a networked course, *British Journal of Educational Technology*, 33(5), 603–618.
- Mackenzie, D. (1999) Recent developments in the triartite interactive assessment delivery system (TRIADS), *Proceedings of the 3rd Annual Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Mackenzie, D., Hallam, B., Baggott, G. & Potts, J. (2002) TRIADS experiences and developments. A panel discussion, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Mackenzie, D. & O'Hare, D. (2002) Empirical prediction of the measurement scale and base level 'Guess Factor' for advanced computer-based assessment, *Proceedings of the 6th International Conference of Computer Aided Assessment* (Loughborough, Loughborough University), 187–201.
- Mackenzie, D., O'Hare, D., Paul, C., Boyle, A., Edwards, D., Willimas, D. & Wilkins, H. (2004) Assessment for learning: the Triads assessment of learning outcomes project and the development of a pedagogically friendly computer-based assessment system, in: D. O'Hare & D. Mackenzie (Eds) *Advances in computer aided assessment* (Birmingham, SEDA), 11–25.
- Mason, S. (2003) Electronic security is a continuous process, *Computer Fraud and Security*, 2003(1), 13–15.
- Maughan, S., Peet, D. & Willmot, A. (2001) On-line formative assessment item banking and learning support, *Proceedings of the 5th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Mayer, R. E. (2002) A taxonomy for computer-based assessment of problem solving, *Computers in Human Behaviour*, 18(6), 623–632.
- Mayer, D. K., Sims, V. K. & Koonce, J. M. (2001) Comprehension and workload differences for VDT and paper-based reading, *International Journal of Industrial Ergonomics*, 28(6), 367–378.
- McAlpine, M. (2002) *Principles of assessment* (Luton, CAA Centre).
- McCabe, M. & Barrett, D. (2003) CAA scoring strategies for partial credit and confidence levels, *Proceedings of the 7th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 209–219.
- McKenna, C. & Bull, J. (2000) Quality assurance of computer-assisted assessment: practical and strategic issues, *Quality Assurance in Education*, 8(1), 24–31.
- McLaughlin, P. J., Fowell, S. L., Dangerfield, P. H., Newton, D. J. & Perry, S. E. (2004) Development of computerised assessment (TRIADS) in an undergraduate medical school, in: D. O'Hare & D. Mackenzie (Eds) *Advances in computer aided assessment* (Birmingham, SEDA), 25–32.
- Mills, C. N., Potenza, M. T., Fremer, J. J. & Ward, W. C. (2002) *Computer-based testing. Building the foundation for future assessments* (Mahwah, Lawrence Erlbaum Associates).
- Morgan, M. R. J. (1979) MCQ: an interactive computer program for multiple-choice self testing, *Biochemical Education*, 7(3), 67–69.
- Nugent, G. (2003) On-line multimedia assessment for K-4 students, *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications* (Hawaii, AACE), 1051–1057.
- O'Leary, R. & Cook, J. (2001) Wading through treacle: CAA at the University of Bristol, *Proceedings of the 5th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Pain, D. & Le Heron, J. (2003) WebCT and online assessment: the best thing since SOAP? *Educational Technology and Society*, 6(2), 62–71.

- Paterson, J. S. (2002) Linking on-line assessment in mathematics to cognitive skills, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 295–306.
- Paxton, M. (2000) A linguistic perspective on multiple choice questioning, *Assessment and Evaluation in Higher Education*, 25(2), 109–119.
- Phipps, L. & McCarthy, D. (2001) Computer assisted assessment and disabilities, *Proceedings of the 5th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Pollock, M. J., Whittington, C. D. & Doughty, G. F. (2000) Evaluating the costs and benefits of Changing to CAA, *Proceedings of the 4th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Powers, D. E., Burstein, J. C., Chodorow, M., Fowles, M. E. & Kukich, K. (2002) Stumping e-rator: challenging the validity of automated essay scoring, *Computers in Human Behaviour*, 18(2), 103–134.
- Race, P. (1995) The art of assessing, *The New Academic*, 4(3).
- Reid, N. (2002) Designing online quiz questions to assess a range of cognitive skills, *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications* (Denver, AACE), 1625–1630.
- Ricketts, C. & Wilks, S. (2002a) What factors affect students opinions of computer-assisted assessment? *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 307–316.
- Ricketts, C. & Wilks, S. (2002b) Improving student performance through computer-based assessment: insights from recent research, *Assessment and Evaluation in Higher Education*, 27(5), 475–479.
- Sabar, N. (2002) Towards principled practice in evaluation: learning from instructors' dilemmas in evaluating graduate students, *Studies in Educational Evaluation*, 28(4), 329–345.
- Sambell, K., Sambell, A. & Sexton, G. (1999) Students' perception of the learning benefits of computer-assisted assessment: a case study in electronic engineering, in: S. Brown, J. Bull & P. Race (Eds) *Computer-assisted assessment in higher education* (Birmingham, SEDA), 179–191.
- Schenkman, B., Fukuda, T. & Persson, B. (1999) Glare from monitors measured with subjective scales and eye movements, *Displays*, 20, 11–21.
- Sclater, N., Davis, H. C., White, S. A., Conole, G. C. & Danson, M. (2003) Technologies for online interoperable assessment, *Proceedings of the CAL03* (Belfast, Elsevier).
- Sclater, N. & Howie, K. (2003) User requirements of the 'ultimate' online assessment engine, *Computers and Education*, 40(3), 285–306.
- Sealey, C., Humphries, P. & Reppert, D. (2003) At the coal face: experience of computer based exams, *Proceedings of the 7th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 357–376.
- SENDA (2001), Special Educational Needs and Disability Act 2001, <http://www.hmso.gov.uk/acts/acts2001/20010010.htm> (Accessed 10 March 2004)
- Sim, G., Malik, N. A. & Holifield, P. (2003) Strategies for large-scale assessment: an institutional analysis of research and practice in a virtual university, *Proceedings of the 7th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 379–390.
- Stephens, D. (1994) Using computer-assisted assessment: time saver or sophisticated distractor? *Active Learning*, 1, 11–15.
- Stephens, D., Bull, J. & Wade, W. (1998) Computer-assisted assessment: suggested guidelines for an institutional strategy, *Assessment and Evaluation in Higher Education*, 23(3), 283–294.
- Stephens, D. & Mascia, J. (1997) *Results of a survey into the use of computer-assisted assessment in institutions of higher education* (Loughborough University). Available online at: <http://www.lboro.ac.uk/service/lttd/flicaa/downloads/survey.pdf> (accessed 18 June 2004).

- Stevenson, A., Sweeny, P., Greenan, K. & Alexander, S. (2002) Integrating CAA within the University of Ulster, *Proceedings of the 6th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 329–340.
- Tannenbaum, R. S. (1999) *Theoretical foundations of multimedia* (New York, W.H. Freeman).
- Thomas, P., Price, B., Paine, C. & Richards, M. (2002) Remote electronic examinations: student experience, *British Journal of Educational Technology*, 33(5), 537–549.
- Valenti, S., Cucchiarelli, A. & Panti, M. (2002) Computer based assessment systems evaluation via the ISO90126 quality model, *Journal of Information Technology Education*, 1(3), 157–175.
- Walker, D. M. & Thompson, J. S. (2001) A note on multiple choice exams, with respect to students' risk preference and confidence, *Assessment and Evaluation in Higher Education*, 26(3), 261–267.
- Warburton, B. & Conole, G. (2003) CAA in UK HEIs—the state of the art? *Proceedings of the 7th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 433–441.
- White, S. & Davis, H. C. (2000) Creating large scale test banks: a briefing for participative discussion and agendas, *Proceedings of the 4th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University).
- Whittington, D. (1999) Technical and security issues, in: S. Brown, J. Bull & P. Race (Eds) *Computer assisted assessment in higher education* (Birmingham, SEDA), 21–28.
- Wiles, K. (2002) Accessibility and computer-based assessment: a whole new set of issues? in: L. Phipps, A. Sutherland & J. Seale (Eds) *Access all areas: disability, technology and learning* (Oxford and York, ALT/TechDis), 61–66.
- Wiles, K. & Ball, S. (2003) Constructing accessible CBA: minor works or major renovations? *Proceedings of the 7th International Computer Assisted Assessment Conference* (Loughborough, Loughborough University), 445–451.
- Wiltfelt, C., Philipsen, P. E. & Kaiser, B. (2002) Chat as media in exams, *Education and Information Technologies*, 7(4), 343–349.
- Wood, D. A. (1960) *Test construction* (Columbus, Charles E. Merrill Books).
- Yorke, M., Barnett, G., Bridges, P., Evanson, P., Haines, C., Jenkins, D., Knight, P., Scurry, D., Stowell, M. & Woolf, H. (2002) Does grading method influence honours degree classification? *Assessment and Evaluation in Higher Education*, 27(3), 269–279.
- Zakrzewski, S. & Steven, S. (2003) Computer-based assessment: quality assurance issues, the hub of the wheel, *Assessment and Evaluation in Higher Education*, 28(6), 609–623.