IJITCE

ISSN 2347- 3657

International Journal of
Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume2,Issue 3,July 2013

A review of computer-assisted assessment

Jennison V. Asuncion

Abstract

More frequent assessments are being demanded as a consequence of pressure for improved measurement of specified learning outcomes. It is believed that available resources are either staying the same or decreasing, whereas ICT may automate evaluation duties, hence increasing productivity. In this article, we take a look back at CAA and look forward to what the field has in store for the future. In order to track the evolution of CAA from its inception to its widespread application in higher education, a literature search was carried out pertaining to CAA during the last decade. On the one hand, individual academics face challenges including a lack of resources, lethargy, and a risk inclination; on the other hand, institutional hurdles like inadequate funding and cultural considerations are more significant than technical ones.

Introduction

Assessment is a critical catalyst for student learning (for example, Brown *et al.*, 1997) and there is considerable pressure on higher education institutions to measure learning outcomes more formally (Farrer, 2002; Laurillard, 2002). This has been interpreted as a demand for more frequent assessment. The potential for Information and Communications Technology (ICT) to automate aspects of learn- ing and teaching is widely acknowledged, although promised productivity benefits have been slow to appear (Conole, 2004a; Conole & Dyke, 2004). Computer- assisted assessment (CAA) has considerable potential both to ease assessment load and provide innovative and powerful modes of assessment (Brown *et al.*, 1997; Bull & McKenna, 2004), and as the use of ICT increases there may be 'inherent difficul- ties in teaching and learning online and assessing on paper' (Bull, 2001; Bennett, 2002a).

This review of CAA sits within a context of increased ICT use, student diversity, financial constraints and the shift from quality assurance to quality enhancement. The review presents a fresh look at the topic by describing key features of CAA identifying issues and highlighting potential areas for research. It describes progress made in identifying and addressing critical factors associated with implementing CAA.

,Department of information technology *University of Southampton, UK*

International Journal of Information Technology & Computer Engineering
ISSN 2347–3657
Volume2,Issue 3,July 2013

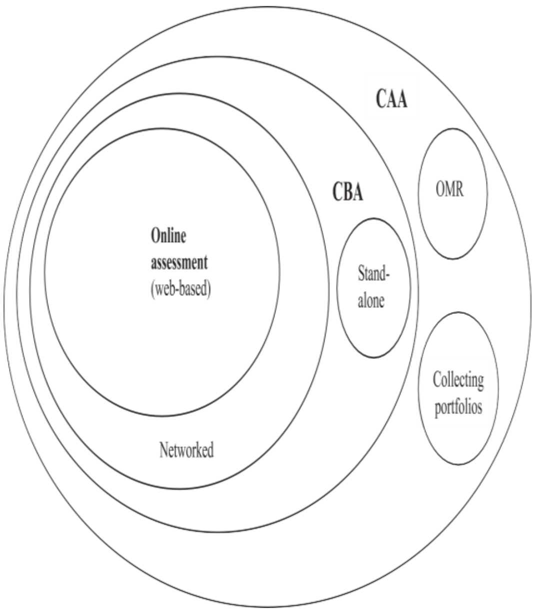

Figure 1. Different types of CAA

CAA in context

The shift towards online testing is well documented (for example, Bennett, 2002a) and different forms of CAA are illustrated in Figure 1. Bull and McKenna recently defined CAA as 'the use of computers for assessing student learning' (2004). Computer-based assessment involves a computer program marking answers that were entered directly into a computer, whereas optical mark reading uses a computer to mark scripts originally composed on paper. Portfolio collection is the use of a computer to collect scripts or written work. Computer-based assessment can be subdivided into stand-alone applications that only require a single computer, appli- cations that work on private networks and those that are designed to be delivered across public networks such as the web (online assessment).

Six ways in which the strategic application of a learning technology such as CAA may add value to the efficiency and effectiveness of the learning process have been identified, along with six factors that may adversely influence it (ALT, 2003). The issues around CAA are similar to those identified for other learning technologies in terms of design and delivery and associated support needs (Seale, 2003). CAA has obvious similarities with the development of Managed Learning Environments in terms of the encountered difficulty of institutional implementation and wide-scale application (Sommerlad et al., 1999). When compared to other learning technologies, the stakes are much greater with CAA, especially when it comes to test usage (QAA, 1998).

The results of CAA assessments may be subject to scrutiny because to the objective tests' explicit scoring methods, which can lead to deficiencies in practice. This, in turn, suggests that risk analysis and management measures are needed, since students are seen as becoming more litigious (Baty, 2004; Zakrzewski & Steven, 2000).

Technical Approach

In order to track CAA's evolution from its inception of widespread usage, a literature search was carried out covering the last ten years. Inclusion was determined by whether or not the content made direct or indirect reference to large-scale CAA review or implementation. The literature would not have been possible without include some previous work, which serves as its basis. Electronic literature indexes, catalogs, and search engines were queried using search terms. This evaluation follows a narrative format that follows the steps taken to overcome operational and cultural challenges to CAA implementation and to identify crucial success factors. This study examines research efforts pertaining to the creation, administration, and evaluation of online assessments, and it takes a look at the literature reviews in light of CAA advancements. It examines emerging patterns of adoption and the related implementation challenges, as well as potential facilitators and impediments to CAA adoption.

Evaluation by category

There are four parts to a multiple-choice question: the question's stem, the possibilities, the right answers, and any distractions (Figure 2). Collections of objects, often from item banks, that are subject-specific make up tests (Sclater, 2004). A wide range of question formats and feedback mechanisms are available, such as multiple-choice, multiple-response, hotspot, matching, ranking, drag-and-drop, multiple-step, and open-ended questions, as well as objective testing with automatic feedback, model answers, annotated tests, and mixed-mode assessments with teacher intervention.Summative assessments are used to grade students' work, whereas formative assessments are used to provide feedback that may be used to improve students' learning. Tutors utilize diagnostic assessments to find out what their students know already, while students use self-assessment to think about what they've learned (O'Reilly & Morgan, 1999; Bull & McKenna, 2004). Other classifications include final/continuous (at the conclusion of a course or throughout) and formal/informal (invigilated or not). According to Sclater and Howie (2003), there are six distinct uses of CAA in education. These include ongoing assessment, diagnostic tests that measure students' knowledge before the course to determine the effectiveness of the teaching, high-stakes summative tests, and authenticated or anonymous self-assessment.Knowledge recall is at the base of Bloom et al.'s (1956) taxonomy of cognitive learning outcomes, which progresses from comprehension to application, analysis, synthesis, and assessment. Some people already

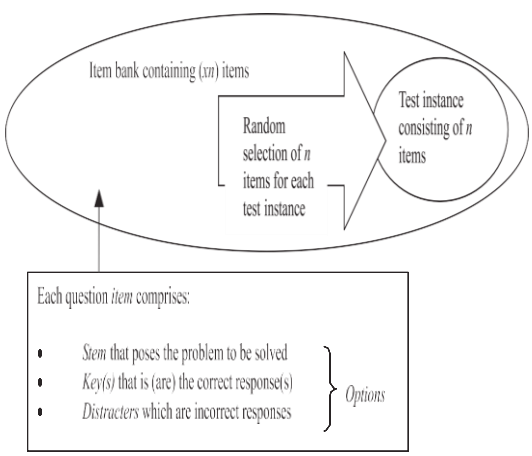

Figure 2. Test selection from an item bank

adopted or expanded upon this (Krathwohl et al., 1964; Imrie, 1995; Anderson et al., 2001). Many believe that universities should focus on cultivating advanced abilities, such as evaluation and critical thinking, even if college instructors evaluate a wide variety of results (Miller et al., 1998, p. 56). (Barnett, 1997). According to McAlpine (2002a), lower-level outcomes in Bloom's taxonomy are often evaluated using a convergent approach, whereby there is only one 'right' answer, but higher-order outcomes are best evaluated using a divergent approach, wherein a variety of informed replies and analyses are allowed. Objective items make it easy to build convergent tests, while divergent assessments have always used lengthier written responses or essays, which are difficult to automatically grade (for example,

instance, see Mason and Grove-Stephensen (2002).Ongoing initiatives in the realm of research and developmentA helpful summary of CAA is given by Bull and McKenna (2004), the fundamentals of evaluation are laid out by McAlpine (2002a), and some of the present research objectives are described by Mills et al. (2002).There are three main categories into which CAA R&D efforts fall. From the creation of standalone components to the detailed description of whole CAA systems, the first one is all about design. The second one is about getting things done and handed off. Analysis, test scoring, and the creation of suitable reporting systems are the main topics of the third.

Design

Product creationCAA enables the incorporation of audiovisual elements and more intricate interactions between items, allowing for more complicated item kinds than paper-based evaluations.computer and student. The evidence suggests that it is not suitable to simply transfer questions from paper-based examinations to online ones; instead, we should reevaluate the questions' formulation in light of their intended purpose. Since paper-based questions are also part of the CAA question-making process, this leads to some very basic questions about their nature. In addition to helping students and instructors comprehend the learning process and question-solving strategies, using stages for certain types of questions (such as mathematical examples) has been beneficial for research and practice (Ashton et al., 2003). The way CAA software

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume2,Issue 3,July 2013

systems record and report on student interactions is called into question by this. Richer data on the pupils is accessible via the reporting methods offered under CAA systems compared to paper-based assessments. A number of significant educational considerations follow from the evolution of online evaluations.

How an evaluation is structured influences its validity, reliability, and the results that students get. There could be a number of ways in which paper and online tests vary. Ward et al. (1980), Outtz (1998), and Fiddes et al. (2002) are just a few examples of studies that have contrasted paper-based and computer-based examinations to investigate this. Specifically, a large-scale examination of Scottish schools and colleges covering a variety of subjects and levels was carried out by the Pass-IT project (Ashton et al., 2003, 2004). Results differ depending on item category, level, and topic area. Mode impact may be caused by several factors, such as examinee qualities, item type, item ordering, local item reliance, and student test-taking experience. Also, people may use various test-taking tactics and have varied cognitive abilities depending on the mode. In order to address these concerns and provide standards for creating suitable administrative processes or statistically modifying item attributes, it is necessary to have a firm grasp of these topics throughout item development.

Researchers are looking at ways to automate assessment as marking objective test scripts is a straightforward repetitive activity, unlike marking essays. Standardized testing in schools, universities, professional entrance exams, and psychological evaluations are all areas where objective testing has become well-established in the US and elsewhere (Bennett, 2002b; Hambrick, 2002).

Problems with item type constraints persist. The validity of using multiple-choice questions (MCQs) to evaluate students' mastery of more complex concepts is an important issue in the realm of objective testing (Pritchett, 1999; Davies, 2002). This view is shared by academics and quality assurance professionals (Bull, 1999; Warburton & Conole, 2003). Assuming they are well-designed, item-based tests have the potential to be useful for gauging a wide variety of student learning outcomes in both undergraduate and graduate programs (Farthing & McPhee, 1999; Duke-Williams & King, 2001). Although there is consistent push for the use of "more complex" question forms, the most common ones are multiple-choice and multiple-response questions (Boyle et al., 2002; Warburton & Conole, 2003) (Davies, 2001).

The creation of digital goods is also a topic of active research and development (Mills et al., 2002). Part of this process involves making item templates that are accurate enough for the computer to make many objects at once without human intervention.precisely measured. For example, compared to more advanced content domain regions, lower-level mathematics is supposedly simpler to reproduce.

Storage facilities for items

Item banks are compilations of questions that are often created in a collaborative effort throughout a certain field. These questions may be categorized based on complexity, skill type, or subject matter. A comprehensive evaluation of item bank developments in the UK was recently carried out by Sclater and colleagues (Sclater, 2004). This assessment included metadata, security, access and authentication, and legal concerns. While Sclater sees item banks as CAA's primary motivator, McAlpine (2002b) contends that students should routinely use item banks to address concerns around authenticity, fairness, security, and quality assurance. Mills et al. (2002) noted that other scholars have investigated the creation, upkeep, and modification of item banks. Security is of the utmost importance when it comes to high-stakes exams; one way to discourage cheating is to employ bigger pools. Research and development efforts aimed at limiting item exposure and ensuring the safety of item banks are expected to continue.

Computer-assisted evaluation

Through computer-adaptive testing (CAT), objective tasks are designed to probe the boundaries of a participant's abilities. The computer-administered exam (CAT) uses the test-taker's past answers to determine the difficulty level of the questions. With this system, students' abilities may be quickly and accurately assessed by raising their ability estimate and presenting more challenging questions in response to right answers, and vice versa. This allows for a succinct testing of several skill levels. To illustrate this point, Lilley and

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume2,Issue 3,July 2013

Barker (2003) used Item Response Theory to build a database of 119 items that had been peer-reviewed. They then administered both "conventional" (non-adaptive) and CAT exams to 133 students. Students did not perceive the CAT exam to be unfair, and their scores were highly congruent with those of the conventional version. Results from CATs may be more accurate and trustworthy than those from more conventional exams as the questions are designed to measure specific levels of competence.

CAD/CAM software platforms

Software solutions for CAA varies in price, adaptability, and scalability, and also differ in the ways in which they facilitate the creation, distribution, and evaluation of online tests. A free system that complies with Question and Test Interoperability (QTI) standards was developed by the TOIA and is part of their suite of assessment management and reporting tools (TOIA, 2004). According to McKenzie et al. (2002), TRIADS offers many question types in various formats to make it easier to measure higher-order learning abilities. In terms of both the original expenditure and the maintenance agreements, commercial products such as Perception (Question Mark Computing Ltd, 2004) are substantial investments. Changes in CAA systems

differ greatly in the control administrators have over the scheduling of exams and the amount of question types provided. Among them, Perception offers support for 18 different sorts of objective items, TOIA for nine (but the commercial edition offers many more), and Blackboard for six. Even though virtual learning environments' assessment tools aren't always the most advanced, they've been crucial in encouraging practitioners to use CAA. The scalability remains an issue, despite the fact that more expensive systems are sometimes peddled with the promise of equal flexibility and devoted support (Danson et al., 2001; Cosemans et al., 2002; Stevenson et al., 2002; Harwood, 2004). The amount of concurrent tests that different programs allow varies widely (Question Mark Computing Ltd, 2004).While early CAA systems operated independently, modern systems are either browser-based or linked via a private network. Although multiple-choice questions continue to make up the majority of CAA exams (Warburton & Conole, 2003), there has been a need for more adaptable question types, such as those that might be challenging or impossible to write down (Bull & Hesketh, 2001; Davies, 2001; Bennett, 2002b). The fact that CAA software is beginning to accommodate more and more question kinds, including 'new' question types, is indicative of this trend. The goal of vendors is to maximize the number of item kinds that their product supports. Although many are only variations on fundamental question kinds, which makes comparing CAA products difficult, customers consider this as a key criterion of flexibility and a way for providers to distinguish themselves (Paterson, 2002).

Cooperation between

The capacity to move queries across systems relies on interoperability. Since practitioners are using a variety of ICT tools to supplement their instruction, they may find it convenient to create questions on one platform and administer exams in another. How fully interoperable are present CAA systems is a matter of debate (Sclater et al., 2002).Two further justifications for interoperability are laid down by Lay and Sclater (2001). We need to know, first, whether the item banks will still be used when the CAA systems go out of service, and second, if we can move student test scores to our school's database. Making sure that users can keep using their old questions and exams even when they switch institutions or use other CAA systems is another major motivator for interoperability. While the QTI standard developed by the IMS (2003) Consortium serves as a good foundation, further research is obviously required (Sclater et al., 2002; Sclater & Howie, 2003).

DeliveryConformity with CAA practice standards as publishedThe growing use of CAA "has elevated," according to the most current code of practice for the use of IT in assessment delivery (BS 7988: 2002).

problems with the reliability and validity of IT-based evaluations, leading to a patchwork of methods (BSI, 2002, p. 11). Its purpose is to prove CAA's legitimacy, authenticity, security, and fairness in order to raise its profile and promote its use. The code's emphasis on CAA test delivery, however, may cause earlier phases of assessment preparation and quality assurance to be under-emphasized. 'A bad evaluation, provided adequately, would [still] adhere to BS,' argue Boyle

ISSN 2347–3657
Volume2,Issue 3,July 2013

and O'Hare (2003, p. 72). More suitable standards for practice were found to be the following: the Association of Test Publishers' Guidelines for Computer-based Testing, the American Educational Research Association's Standards for Educational and Psychological Testing (2003), and the Scots Qualifications Authority's Guidelines for Online Assessment in Further Education (2003).

Extending CAA

Individuals generally use CAA on an as-needed basis without a comprehensive plan or institutional IT infrastructure, which hinders its implementation. Bull (2001) and Boyle and O'Hare (2003) found that this might impede or postpone embedding. If academics and support workers are to keep up with the rapid speed of technical and software advancements, "retooling is an issue which affects on research and development," according to Bull (2001, p. 11). Practitioner isolation and underfunding are two potential downsides of small-scale development, whereas potential upsides include practitioners having control over the process (Kennedy, 1998). Danson et al. (2001), Cosemans et al. (2002), and Warbur- ton and Harwood (2004) all note that CAA implementation at higher education institutions has unique risks and advantages. In order to scale up for full-scale institutional deployment, which encompasses all potential uses, it seems that cultural rather than technological obstacles will need to be resolved. The difficulties linked with the adoption of other forms of educational technology are similar to those that Bull (2001, p. 11) notes, saying that "the organisational and pedagogical issues and challenges surrounding the take-up of CAA often outweigh the technical limitations of software and hardware." (See, for instance, Seale, 2004).Concerns over dangers will most certainly persist for individual practitioners (Harwood & Warburton, 2004). Incremental technological advancements may be able to overcome operational hurdles, but cultural ones are showing to be more resilient.

Important elements controlling the adoption of CAA

By this point, the conventional wisdom about evaluation methods has been mostly dispelled. Traditional evaluations do not always function well, nevertheless (see, for instance, Goddard, 2002). A lot of the same barriers and facilitators that are important for conventional evaluation also apply to CAA. These long-ago concerns in conventional wisdom have been revived by the advent of CAA. One reason why institutions should prioritize CAA is because it prompts a broader reevaluation of assessment methods (Bull & McKenna, 2004).The variables controlling uptake have been the subject of several investigations. In 1995, the first study of CAA usage in the UK was carried out by Mascia and Stephens utilizing a

a survey with ten questions that received forty-five answers. Among the issues they highlighted were the following: subject-related dependencies, the need for institutional support in the form of training and resources, sufficient time to produce CAA assessments, and the integration of CAA into current assessment processes. Tools familiarity, well-planned processes to resolve reliability and security concerns, and support staff participation were critical operational considerations (Stephens & Mascia, 1997).

After a span of four years, the CAA Centre surveyed colleges and universities throughout the country to gauge use and sentiment. Building on the work of Stephens and Mascia (Bull, 1999), this survey included twice as many items, with many things requiring several parts to complete. Researchers, quality assurance specialists, and staff developers filled out over seven hundred fifty surveys (McKenna, 2001). Fifty people, largely academic CAA fans, participated in an online revamp of the 1999 survey that Warburton and Conole (2003) tested.

Cost, including both individual effort and the price of commercial software, was considered the most institutional obstacle. Also mentioned were unrealistic expectations, inherent conservatism, and a dearth of instructional and technical resources. Integration, security, and copyright concerns with Managed Learning Environments were less of a worry to respondents. People also thought there was a high learning curve to the technology and making their own unique kinds of CAA questions. See also Boyle and O'Hare (2003) for more on the topic of the challenges of developing objective questions that accurately measure learning outcomes at the university level. Less often mentioned were issues

ISSN 2347–3657
Volume2,Issue 3,July 2013

of lack of support, cultural opposition, and technophobia, suggesting a credibility gap between the claims made by CAA proponents and what respondents believed could be achieved. We also discussed related topics such as academics working alone, individual inertia, and usability. Large-scale adoption of CAA was seen as being driven by subject-specific shared question banks and the importance of exemplars, rather than by the availability of "evangelists" or adherence to institutional rules.

Although academic dedication was mentioned as a key facilitator, faculty support for CAA seems to be minimal, and external financing is the main source of support for CAA at this level. The necessity to include CAA into regular lessons was another critical component. The incorporation of multimedia and effective interoperability were also mentioned. This was especially true between CAA systems and Virtual Learning Environments. While the majority of systems were web-based, a tiny fraction made use of optical mark reading, and a large number of respondents provided CAA over restricted networks. Even while certain summative CAA assessments were worth up to 100%, only a third were under supervision, and the majority of those tests only counted for a third or less of the total score. Using CAA, we administered tests to classes ranging in size from a few kids to over 200. Bull (1999) and Warburton and Conole (2003) found that there were clear subject-specific variations in the absorption of CAA. Notably, quality assurance staff members could not find many enabling circumstances, which may be a reflection of their generally poor view of CAA (Bull, 1999; Bull & McKenna, 2000; Bull & Hesketh, 2001).In the 1999 poll, 90% of the obstacles and 65% of the facilitators were deemed cultural, highlighting the paramount importance of cultural concerns. Research by Hambrick (2002) onAmong the many variables influencing the widespread use of formal online assessment in the United States's K-12 education system, two categories stand out: operational and cultural. Cultural influences accounted for one-third of the components found in a rigorous risk assessment undertaken by Zakrzewski and Steven (2000).

Analysis

Assessing and rating itemsOne advantage of CAA is that it allows for the recording and analysis of student interactions, which may give a more comprehensive picture of learning. It is possible to evaluate the performance of specific questions or students using a number of different approaches. Then, ineffective things may be removed or instructional approaches changed. There are several applications for automatically recorded data, such as investigating the correlation between response speed and accuracy. Be cautious when interpreting findings, however, since students' wrong answers could reveal a deeper level of comprehension than meets the eye. For instance, if a student gets a grammar question wrong on a foreign language exam, it might be because they have a more advanced cognitive knowledge of the subject. While all examinees, regardless of ability level, took more time to process more complex things, those with lower abilities spent more time encoding the stem and those with higher abilities spent more time processing what came after. This finding was reported by Gitomer et al. (1987).It is essential that assessments be dependable and genuine. The fact that CAA provides uniform labeling is one of its benefits. Scoring may be done in a variety of ways, from just giving a mark to the right answer to more complex approaches that include negative scoring and variable scoring. Classical Test Theory and Latent Trait Analysis (LTA) are the two primary approaches to item statistics, alongside Rasch analysis and Item Response Theory. The former is less complex and uses a test-level evaluation, whereas the latter examines problems individually. Please refer to McAlpine, 2002c, and Mills et al., 2002 for more information on these topics. The use of Classical Test Theory, Item Response Theory, and Rasch analysis is investigated by Boyle et al. (2002) using a set of 25 questions administered to 350 test students. Specifically for high-stakes assessment, they found that many practitioners' current approach to CAA—that is, ignoring the rigorous quality assurance of items—is unworkable. Staff members engaged in creating CAA tests should be required to undergo training in item building and analysis, according to Boyle and O'Hare (2003). Additionally, items should be peer-reviewed and piloted before being used. Curriculum development and quality assurance may both benefit from statistical analysis (Bull & McKenna, 2004). Encouraging students to reflect on their strengths and areas for improvement, as well as

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657
Volume2,Issue 3,July 2013

compare their progress to predetermined goals, is the goal of this sort of feedback.Two primary areas deal with worries that pupils could predict the answers.methods: first, by reducing the weight of guesswork on tests; and second, by shifting from a basic tariff system in which "one right answer equals one point" to one that incorporates negative marking. According to research by Davies (2002), Gardner-Medwin & Gahan (2003), and McAlpine & Hesketh (2003), what is known as confidence-based evaluation is when the marks provided for an answer are projected to be right.

Strategies such as interleaving participants taking separate tests, randomizing question and answer order, presenting "blinker screens," and adequate invigilation have been proposed to alleviate concerns about cheating in summative assessments (BSI, 2002; Bull & McKenna, 2004; Pain & LeHeron, 2003). The usage of item banks is one of the additional strategies.

The way forward

Investigation on the disparities between online and paper-based evaluations in a variety of grade levels and topic areas is ongoing. We also need to figure out how to put this information to use in order to make better questions and provide assessments in the right way. The relationship between effective question design and learning outcomes also need more investigation. For the foreseeable future, it seems that CAA will remain in demand specialized knowledge, especially in the realm of product development. Graphical, animated (FLASH), and open-ended JAVA item types offered by more current CAA systems suggest that concerns regarding cost-benefit may be partly addressed by integrating additional technologies, such as multimedia (TOIA, 2004; Question Mark Computing Ltd, 2004).The difficulty of evaluating results at a higher level may be alleviated by using automated essay marking. Unfortunately, this requires a substantial investment of time and expertise, and therefore is not yet included in many CAA systems (Christie, 2003).The development of new standards for test interoperability beyond items may be prompted by the need for greater interoperability in system design. Questionmark Markup Language and other de facto standards are being replaced by more open ones, such as the QTI specification by the IMS Consortium (IMS, 2003). One may argue that the present "off the shelf" commercial CAA systems do not meet the requirements of institutions, in contrast to the "ideal" CAA system proposed by Sclater and Howie (2003). Six functional roles (question authors, viewers, and validators; test authors, viewers, and validators); three related to the sale of items or assessments; three administrative roles; six related to processing results; and three related to delivery (test-taker, timetabler, and invigilator) are identified as possible roles that a full-scale CAA system might require.

In summary,

Current actions in the design, delivery, and analysis of online assessment have been reviewed in this work. Similar to other learning technologies, CAA has both opportunities and challenges in terms of adoption and usage; nevertheless, due to its key role in education and the perceived high risk it poses, CAA stands apart. In particular, the study poses difficult pedagogical problems about the function of assessment and its connection to learning, which in turn raises basic concerns regarding the whole teaching and learning process.We need to create new models to investigate the emerging role of technology and its potential effects on evaluation. It is also important for us to comprehendlearn more about the ways to overcome these obstacles and make the most of these resources. Because of this, there are serious problems with our expectations for and goals for kids' education. Furthermore, there is a widespread movement in education to evaluate learning processes rather than outcomes. Because of this, we should think about what we want to evaluate and the best way to accomplish so. Similarly, problems emerge as a result of the new ways of communicating and working together. In what ways may the quality of student learning be gauged from the exchanges that take place in online discussion forums? Are "the lurkers," or those who don't actively participate, choosing to stay on the sidelines or finding other ways to learn (such as by observing and contemplating the posts of others)? Exploring whether new types of assessment could emerge as a consequence of the effect of technologies is one of several problems listed by Conole (2004b) that CAA researchers need to address. To reach our aim of making the most of the opportunities presented by technology to enhance education, assessment, and learning, we must investigate and resolve the basic concerns raised by CAA, which concern the whole learning and teaching process.

References

AERA (1999) *Standards for educational and psychological testing* (Washington, DC, AERA).

ALT (2003) The future of higher education: the Association for Learning Technologies response to the white paper. Available online at: http://www.alt.ac.uk/docs/he_wp_20030429_final.doc (accessed 19 December 2003).

ATP (2003) *Guidelines for computer-based testing* (Washington, DC, ATP).

Anderson, L., Krathwohl, D. & Bloom, B. (2001) *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives* (New York, Longman).

Ashton, H. S., Schofield, D. K. & Woodgar, S. C. (2003) Piloting summative web assessment in secondary education, in: J. Christie (Ed.) *Proceedings of the 7th International Computer–Assisted Assessment Conference,* Loughborough University, 8–9 July 2003.

Ashton, H. S., Beavers, C. E., Schofield, D. K. & Youngson, M. A. (2004) Informative reports—experiences from the Pass-IT project, in: M. Ashby and R. Wilson (Eds) *Proceedings of the 8th International Computer–Assisted Assessment Conference,* Loughborough University, 6–7 July 2004.

Barnett, R. (1997) *Higher education: a critical business* (Buckingham, Society for Research into Higher Education and the Open University Press).

Baty, P. (2004, March 12) Litigation fees top £15m as academic disputes grow, *Times Higher Education Supplement,* 2004, p. 7.

Bennett, R. E. (2002a) Inexorable and inevitable: the continuing story of technology and assessment. *Journal of Technology, Learning and Assessment,* 1. Available online at: http://www.bc.edu/research/intasc/jtla/journal/pdf/v1n1_jtla.pdf (accessed 17 July 2004).

Bennett, R. E. (2002b) *Using electronic assessment to measure student performance.* The State Education Standard, National Association of State Boards of Education.

Bloom, B., Engelhart, M., Furst, E., Hill, W. & Krathwohl, D. (1956) *Taxonomy of educational objectives: the classification of educational goals. Handbook 1: cognitive domain* (London, Longman). Boyle, A., Hutchison, D., O'Hare, D. & Patterson, A. (2002) Item selection and application in higher education, in: M. Danson (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: J. Christie (Ed.) *7th International CAA Conference,* Loughborough University, 8–9 July 2004.

Brown, G., Bull, J. & 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* (London, BSI).

Bull, J. (1999) Update on the National TLTP3 Project The Implementation and Evaluation of Computer-assisted Assessment, in: M. Danson (Ed.), *3rd International CAA Conference,* Loughborough University, 16–17 June 1999.

Bull, J. (2001) *TLTP85 implementation and evaluation of computer-assisted assessment: final report.* Available online at: www.caacentre.ac.uk/dldocs/final_report.pdf (accessed16 March 2003). Bull, J. & Hesketh, I. (2001) Computer-assisted assessment centre update, in M. Danson & C. Eabry (Eds) *5th International CAA Conference,* Loughborough University, 2–3 July 2001. Bull, J. & Mckenna, C. (2000) Computer-assisted assessment centre (TLTP3) update, in: M. Danson (Ed.) *4th International CAA Conference,* Loughborough University, 21–22 June 2000.

Bull, J. & McKenna, C. (2004) *Blueprint for computer-assisted assessment* (London, RoutledgeFalmer) Christie, J. (2003) Automated essay marking for content—does it work?, in J. Christie. (Ed.) *7th International CAA Conference,* Loughborough, 8–9 July 2001.

Conole, G. (2004a) *Report on the effectiveness of tools for e-learning,* report for the JISC commissioned Research Study on the Effectiveness of Resources, Tools and Support Services used by Practitioners in Designing and Delivering E-Learning Activities.

Conole, G. (2004b) *Assessment as a catalyst for innovation,* invited keynote and paper for the Quality Assurance Agency for Higher Education, Assessment Workshop 5. Available online at: http://www.qaa.ac.uk/scottishenhancement/events/default.htm (accessed 10 September 2004).

Conole, G. & Dyke, M. (2004) What are the affordances of Information and communication technologies?, *ALT-J,* 12(2), 111–123.

Cosemans, D., Van Rentergem, L., Verburgh, A. & Wils, A. (2002) Campus wide setup of question mark perception (V2.5) at the Katholieke

Universiteit Leuven (Belgium)—facing a large scale implementation, in: M. Danson (Ed.) *5th International CAA Conference,* University of Loughborough, 2–3 July 2001.

Danson, M. (2003) *Implementing online assessment in an emerging MLE: a generic guidance document with practical examples* (Bristol, JISC).

Danson, M., Dawson, B. & Baseley, T. (2001) Large scale implementation of question mark perception (V2.5)—experiences at Loughborough University, in: M. Danson (Ed.) *6th International CAA Conference,* Loughborough University, 2–3 July 2001.

Davies, P. (2001) CAA must be more than multiple-choice tests for it to be academically credible?, in: M. Danson & C. Eabry (Eds) *5th International CAA Conference,* Loughborough University. Davies, P. (2002) There's no confidence in multiple-choice testing, in: M. Danson (Ed.) *6th International CAA Conference,* Loughborough University, 4–5 July 2002.

Duke-Williams, E. & King, T. (2001) Using computer-aided assessment to test higher level learning outcomes, in: M. Danson & C. Eabry (Eds) *5th International CAA Conference,* Loughborough University, 2–3 July 2001.

Farrer, S. (2002) End short contract outrage, MPs insist, *Times Higher Education Supplement.*

Farthing, D. & McPhee, D. (1999) Multiple choice for honours-level students?, in: M. Danson (Ed.) *3rd International CAA Conference,* Loughborough University, 16–17 June 1999.

Fiddes, D. J., Korabinski, A. A., McGuire, G. R., Youngson, M. A. & McMillan, D. (2002) Are mathematics exam results affected by the mode of delivery?, *ALT-J,* 10(6), 1–9.

Gardner-Medwin, A. & Gahan, M. (2003) Formative and summative confidence-based assess- ment, in: J. Christie (Ed.) *7th International CAA Conference,* Loughborough University, 8–9 July 2003.

Gitomer, D. H., Curtis, M. E., Glaser, R. & Lensky, D. B. (1987) Processing differences as a function of item difficulty in verbal analogy performance, *Journal of Educational Psychology,* 79, 212–219.