Suto, W.M.I., Nádas, R. & Bell, J.F. (2009). Who should mark what? A study of factors affecting marking accuracy in a biology examination. *Research Papers in Education*. (Published online to date).

van Avermaet, E. (1988). Social Influence in Small Groups. In: M. Hewstone, W. Stroebe, J-P. Codol & G. M. Stephenson (Eds.), *Introduction to Social Psychology*. 350–380. Oxford: Basil Blackwell.

Vidal Rodeiro, C. (2007). Agreement between outcomes from different double-marking models. *Research Matters: A Cambridge Assessment Publication*, **4**, 28–34.

Whetton, C. & Newton, P. (2002). *An evaluation of on-line marking*. Paper presented at the 28th International Association for Educational Assessment Conference, Hong Kong SAR, China, September.

Wilmut, J., Wood, R. & Murphy, R. (1996). *Review of Research into the Reliability of Examinations*. A discussion paper prepared for the School Curriculum and Assessment Authority.

Wilmut, J. (1984). A pilot study of the effects of complete or partial removal of marks and comments from scripts before re-marking them. *AEB Research Report RAC315*.

Zimbardo, P. & Leippe, M.R. (1991). *The Psychology of Attitude Change and Social Influence*. New York: McGraw Hill. (cited in Murphy *et al.*, 1995).

NEW TECHNOLOGIES

Why use computer-based assessment in education? A literature review

Matt Haigh Research Division

Introduction

The aim of this literature review is to examine the evidence around the claims made for the shift towards computer-based assessment (CBA) in educational settings. In this examination of the literature a number of unevidenced areas are uncovered, and the resulting discussion provides the basis for suggested further research alongside practical considerations for the application of CBA.

The review looks at academic literature from UK and international contexts, examining studies that are based in educational settings from primary education to higher education. It should be noted that the literature identified predominantly emerges from higher education contexts in the UK.

Background

CBA first emerged in educational settings in the 1950s and has undergone a steady expansion in use. Burkhardt and Pead (2003) provide a useful summary of the development of CBA in educational settings for each decade between 1950 and 2000:

1950s: Early computers offered games, puzzles and 'tests'; compilers were designed to identify errors of syntax, and later of style, in computer programs.

1960s: The creators of learning machines, in which assessment always plays a big part, recognised the value of computers for delivering learning programmes.

1970s: The huge growth of multiple-choice testing in US education enhanced the attractions of automatic marking, in a self-reinforcing cycle.

1980s: A huge variety of educational software was developed to support learning, with less emphasis on assessment.

1990s: Along with the continuing growth of multiple-choice testing, integrated learning systems, a more sophisticated development of the learning machines of the 1960s, began to be taken more seriously.

Since the 1990s, the explosive growth of the internet has begun to raise the possibility that testing online, on-demand might replace the traditional 'examination day' model, although many technical and educational challenges remain.

(Burkhardt and Pead 2003, p.134)

This history highlights the varying degree to which assessment has formed part of technology-facilitated pedagogy, along with the dangers of allowing technology to dictate assessment practices such as with the permeation of multiple-choice testing in the US during the 1970s detailed by Clarke, Madaus, Horn, and Ramos (2000).

The accompanying expansion in research activity can be illustrated by interrogating online-databases and filtering by year of publication as illustrated in Figure 1. This indicates that CBA developments in the mid-1990s, highlighted in the quote above, spawned a dramatic increase in the research literature available.
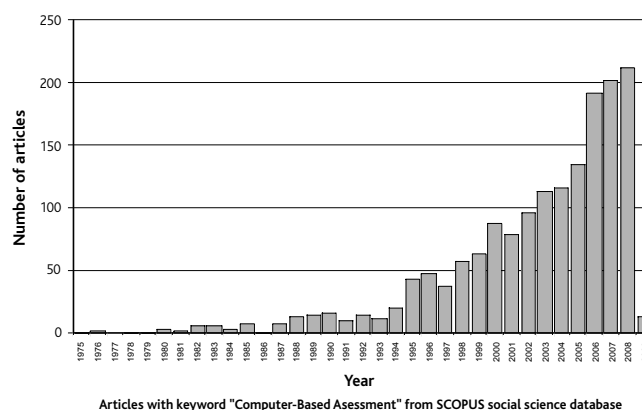

Articles with keyword "Computer-Based Asessment" from SCOPUS social science database

Figure 1: An illustration of CBA research activity

Note that CBA covers a broad range of assessment types, from high-stakes multiple-choice tests through to compilation of assessment evidence in electronic portfolios. This review encompasses this range, however it is quite plausible that the research discussed may only apply to a subset of these assessment types and the reader should consider this caveat throughout.

Strategy for the literature review

In line with approaches to reviewing that make explicit the approach for searching and managing the literature, this section sets out a description of the approach taken. Initially the literature searched emerged from personal professional knowledge. This was then expanded via a number of strategies:

- The use of bibliographic databases and search engines (Scopus, British Education Index, ERIC, Web of Knowledge, Psycinfo, Zetoc, Google Scholar, Directory of Open Access Journals, Education-line, Educational Evidence Portal, Multiverse, Intute);

- Identification of a number of key journals: *British Journal of Educational Technology, Assessment in Education: Principles, Policies and Practice and ALT-J*; a subsequent search of these journal indexes provided additional literature.

- The citation index tools in SCOPUS were also used to identify the most frequent citations in the literature.

Given the number of items of literature emerging from these approaches and the scope of this article, no systematic attempt was made to reference additional repositories of 'grey literature'[1]. All emerging literature was collated and categorised using bibliographic software. The criteria for inclusion were:

- Research carried out in educational institutions and available in the public domain (this excluded work-based training and the use of CBA for recruitment).

- Research included a component of evaluation[2] of the use of CBA.

- A focus on research published post-1995: given the development of technology, particularly the explosion of internet use in the mid-1990s, older studies evaluated different computing technology; therefore pre-1995 studies have only been included as an exception.

It should be noted that a significant proportion of the literature identified was based on case-study methodologies.

Three approaches were taken to extract salient themes from the research:

- Using a set of key questions that were explored using the literature.

- Identifying literature which presents an overview of CBA use, and extracting key themes.

- Use of the tagging system[3] employed to code literature in the bibliographic software.

Figure 2 illustrates this approach.

Overview of the literature emerging from each strategy

1. Using key questions to identify themes

Why use Computer-Based Assessment in Education?

The most immediate claim that emerges from key texts is that CBA is a facilitator of formative assessment (Brown, Race and Bull, 1999). A discussion of the relationship between CBA and formative assessment would seem inevitable given the relentless interest in Black and Wiliam's work encompassed in their publication 'Inside the Black Box' (Black and Wiliam, 1998), aspects of which have made their way into UK Government educational policy (DCSF, 2008). Therefore a further question emerges:
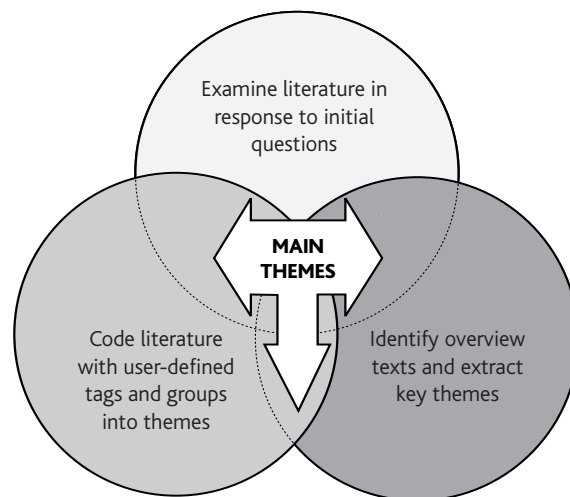

Figure 2: Illustration of the method used to identify themes

What is the relationship between CBA and Formative Assessment?

An examination of the recommended practice in Black and Wiliam's work does indicate areas of formative assessment practice on which CBA might have an impact, for example:

> *Feedback to any pupil should be about the particular qualities of his or her work, with advice on what he or she can do to improve, and should avoid comparisons with other pupils....Tests and homework exercises can be an invaluable guide to learning, but the exercises must be clear and relevant to learning aims. The feedback on them should give each pupil guidance on how to improve, and each must be given opportunity and help to work at the improvement.*

(Black and Wiliam, 1998, p.9)

CBA has the capability to provide feedback for each individual student and, with suitable mechanisms for analysing data, can provide feedback on each student's strengths and weaknesses in relation to their responses to assessment items.

The automated marking element of CBA has the potential to provide timely feedback to enable students to engage in self-assessment. However, feedback from CBA by itself is unlikely to develop the self-assessment skills of students, as Black and Wiliam point out:

> *For formative assessment to be productive, pupils should be trained in self assessment so that they can understand the main purposes of their learning and thereby grasp what they need to do to achieve.*

(Black and Wiliam 1998, p. 10)

It is less clear how CBA may be used in relation to other points raised by Black and Wiliam such as ensuring "The dialogue between pupils and a teacher should be thoughtful, reflective, focused..." (p.12).

The next key question emerges from the discourse in 'Computer Assisted Education in Higher Education' (Brown *et al*., 1999), where the following statement is made:

---

1. Documents not formally published through traditional channels, e.g. government technical reports, commercial product evaluations.

2. Some papers, although set in the context of a CBA environment, were not evaluating the application of CBA per se, but often another aspect of the associated programme.

3. In 'tagging' each piece of literature can be assigned any number of user-defined codes (e.g. 'higher education' 'Formative assessment' 'case-study') which are stored by the bibliographic software alongside the item in question. These tags can then be searched, for example, to find all literature with the 'higher-education' tag associated with it.

Figure 3: Hierarchical tagging system
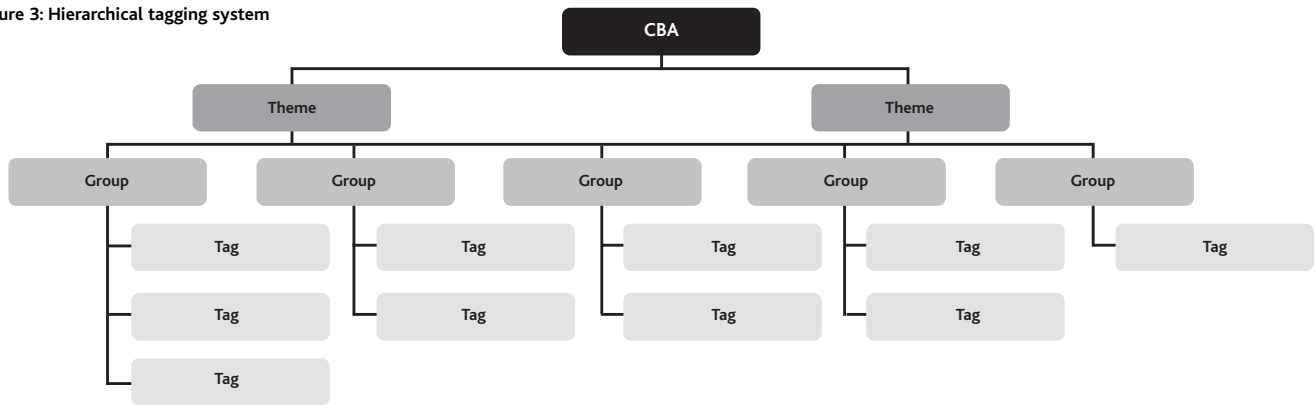
…in most subject disciplines the use of information and communications technologies is expanding rapidly and students are learning a higher proportion of the curriculum using computer-based resources…The gap between how students learn and how they are assessed is widening.

(Brown *et al.*, 1999, p.205)

This provokes the following line of inquiry:

What is the relationship between CBA and students' methods of learning?

The subject of the interrelationships between assessment and learning is much debated. This question will be considered by drawing on Gipps' theory of educational testing (Gipps, 1994), in which the relationship to learning is much discussed: "The implication of work in cognitive science for the assessment of student learning, is that we need to focus on the models that students construct for themselves" (p.29). Therefore, if the models employed by students in their learning are strongly built around a technology-supported environment, then there is a clear argument for the use of CBA in educational assessment.

Gipps also discusses the importance of a wider approach to assessment: "We need a much wider range of assessment strategies to assess a broader body of cognitive aspects than mere subject-matter acquisition." (p.10). The implication for CBA here is that if our 'broader body of cognitive aspects' includes those associated with technology use, then CBA would be the associated assessment strategy.

2. Other claims made from overview texts on the use of CBA

An examination of texts with an overview of CBA derives a number of further claims for CBA. First, there are those who advocate CBA for virtues of efficiency: both Brown *et al.* (1999) and Thelwall (2000) talk of reducing workload by automation; Bull and McKenna (2003) indicate that CBA can be used to decrease marking loads and ease administrative efficiency. In a similar vein Linn, Baker and Dunbar (1991) put forward eight criteria for the evaluation of new assessment types, one of which sits under the heading 'cost and efficiency'.

It is interesting that the notion of efficiency is entering the educational discourse; it could be proposed that this is a managerial function of CBA. However, it is possible to argue that education should be concerned with efficiency: Brown *et al.* (1999) talk of the reduction in resource per student in higher education and the difficulties in extending traditional assessment to meet demand. There is concern that the term 'efficiency' is being used as a cover for a reduction in quality of education, and a justification for the reduction in public-spending on

education (Welch, 1998). This implies that efficiency is directly related to the quality of education, and it is on this basis that the relationship between CBA and efficiency can be an educational issue.

Also emerging from key texts is a theme of motivation: Bull and McKenna, (2003) propose that CBA allows one to increase frequency of assessment to motivate students to learn and encourage students to practice skills. This seems to imply that increased frequency of assessments is a factor in motivating students. This is in contrast to other research indicating that testing is seen to decrease students' motivation to learn (Harlen and Deakin Crick, 2003).

Both Thelwall (2000) and Conole and Warburton (2005) raise the issue of the difficulty of institutional implementation and wide-scale use associated with CBA, however factors that alleviate these difficulties (e.g. the development of staff knowledge of CBA) have also been proposed (Ely, 1999).

Themes emerging from tagging in bibliographic software

As the most relevant literature was collated, the content was coded with user defined tags in the bibliographic software. These codes could then be grouped to identify common elements, which were labelled 'groups'. In a similar process, these 'groups' were assembled into common elements called 'themes'. Figure 3 illustrates this hierarchical scheme of coding. The 'tags' with common concepts are collated into 'groups', which are further collated into themes.

In all, 289 items of literature were examined; Figure 4 illustrates the application of the hierarchy to a set of tags. As an example, the tags 'computer adaptive testing', 'versatility of CBA' and 'transition from paper to screen' have all been put into a group labelled 'new assessment models'. The groups 'new assessment models' and 'criterion validity' have been put together under the theme 'validity'.

Figure 4: Application of the coding hierarchy
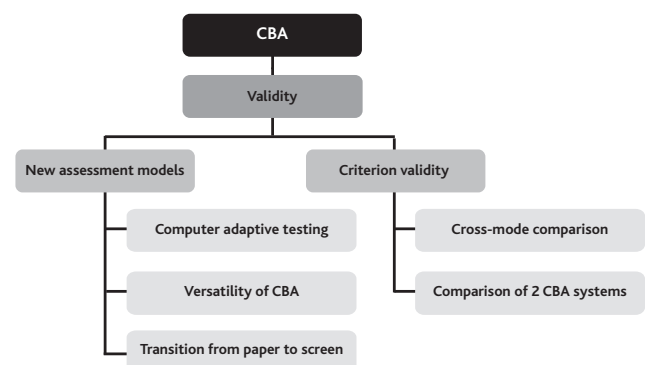
Syntheses of key themes identified in the literature

Using the three strategies (key questions, overview texts and tagging), the analysis of claims made for CBA provides a set of themes that can be seen to converge. The convergence is shown graphically in Figure 5 with the three strategies forming the first column. The second column indicates the key concepts arising from the first two strategies, and a set of groups arising from the tagging strategy. The arrows then show the links to five core themes that emerge across all three strategies.
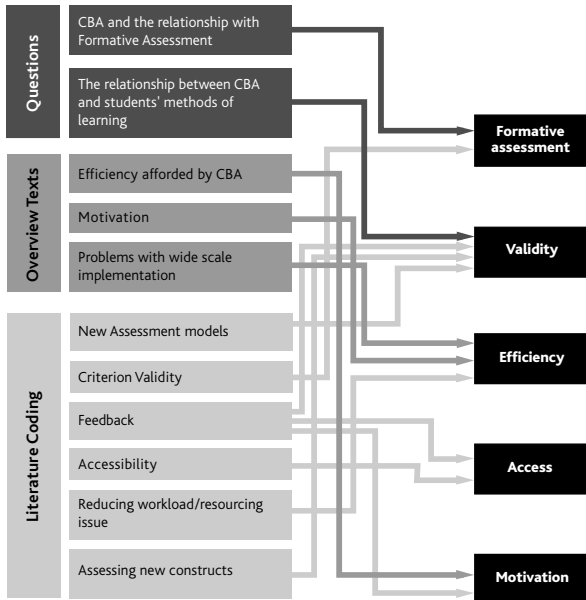
Figure 5: Representation of the emergence of key themes

These five emergent key themes in the existing literature can be more accurately specified as follows:

● Use of CBA to improve the efficiency of assessment programmes (efficiency).

● Use of CBA to facilitate or enhance formative assessment practices (formative assessment).

● The effect of CBA on the validity of assessments (validity).

● Use of CBA to facilitate access to assessments (access).

● The effect of CBA on student motivation (motivation).

It is worth noting at this stage that the themes of efficiency and, to a certain extent, validity indicate a system-centric view of education (pre-occupied with measurement and effective use of resources). The themes of formative assessment, access and motivation indicate a more learner-centred view of education.

Examining evidence in the literature

Evidence for efficiency

It would appear that very little empirical evidence exists that CBA improves efficiency. Loewenberger and Bull (2003) struggled to reach conclusions on the cost-effectiveness aspect of efficiency, but hypothesised that CBA would be more suitable for larger groups. Their report indicates that due to factors associated with immaturity of ICT use and resistance to change:

> …*it becomes extremely difficult to obtain hard data that conclusively demonstrate the cost-effectiveness of CBA. Recommendations and*

> *hypotheses in this study could form the basis for further research. It would seem that the evidence for efficiency is far from well-developed.*

(Loewenberger and Bull, 2003, p.38)

Evidence for CBA facilitating formative assessment

The claims made for CBA facilitating formative assessment are largely derived from a number of evaluations of case studies, so generalisations are difficult to establish unless most cases have the same outcome. In order to provide a framework for exploring the studies relating to formative assessment, the emerging research can be referenced to The Assessment Reform Group's 10 principles for formative assessment:

1. *Assessment for learning should be part of effective planning of teaching and learning.*

2. *Assessment for learning should focus on how students learn.*

3. *Assessment for learning should be recognised as central to classroom practice.*

4. *Assessment for learning should be regarded as a key professional skill for teachers.*

5. *Assessment for learning should be sensitive and constructive because any assessment has an emotional impact.*

6. *Assessment should take account of the importance of learner motivation.*

7. *Assessment for learning should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.*

8. *Learners should receive constructive guidance about how to improve.*

9. *Assessment for learning develops learners' capacity for self-assessment so that they can become reflective and self-managing.*

10. *Assessment for learning should recognise the full range of achievements of all learners.*

(Assessment Reform Group, 2002, p.2)

A number of studies cite the availability of immediate feedback for students as a key benefit in this area (Ashton and Wood, 2006; Bull, Quigley and Mabbott, 2006; Peat and Franklin, 2002). This relates to other research that indicates immediacy of feedback is important in the self-assessment process; however, factors other than immediacy are also shown to be important in Clariana, Ross and Morrison (1991). Despite this, Topping, Samuels and Paul (2007) make a strong case for the educational benefits of timely feedback. This in turn relates to the strand of formative assessment related to self-assessment (principle no.9 above).

Studies also indicated that CBA was able to shed more light on student's difficulties with subject knowledge (Jean, Delozanne, Jacoboni, and Grugeon, 1998) or identify students' methods of learning (Bull *et al.*, 2006). This links to the ideas of students receiving constructive guidance (principle no.7 above).

Studies such as Hunt, Hughes, and Rowe (2002) and Lowry (2005) make claims that improved student performance was related to the formative use of CBA. However, it is difficult to establish attainment gains as a direct result of the use of CBA as the meta-analysis of 23

studies by Fuchs and Fuchs (1986) reported significant attainment gains by those involved in non-CBA formative-assessment based interventions.

The use of CBA was reported to increase dialogue between student and teacher in two case studies (McGuire, 2005; Nicol, 2007), this chimes strongly with one of the key points made by Black and Wiliam (1998):

> *The dialogue between pupils and a teacher should be thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas.*
>
> (Black and Wiliam, 1998, p.12)

Therefore evidence that indicates the use of CBA encourages dialogue is starting to align with the ideals of formative assessment practice.

On the other side of the coin, the studies of the formative use of CBA were also scoured for evidence of any negative impacts associated with their implementations. One recurring theme was the difficulty in demonstrating equivalence between CBA and paper-based formats (Ashton and Wood, 2006; Johnson and Green, 2004).

Another negative impact of CBA was the amount of time taken for both students and teachers to 'learn the system' (Jean *et al*., 1998; McGuire, 2005). This has implications in terms of large scale implementations of CBA as illustrated by Nicol (2007) and the difficulty of institutional implementation and wide-scale use highlighted by Conole and Warburton (2005).

There was also evidence that particular systems were promoting a mechanistic approach and confined to the assessment of lower-order skills (McKenna, 2001). However, other studies such as Ridgway and McCusker (2003) contradict this by implying that CBA facilitates the assessment of higher order skills. Together the studies indicate that the assessment of lower or higher order skills may not be a function of CBA, but the way in which it is used.

It would seem that the key themes emerging from the review of studies linking CBA and formative assessment, in order of prevalence and sufficiency of evidence, are as follows:

- The use of CBA for instant feedback and self-assessment.
- The use of CBA to facilitate anytime-anywhere access to formative assessment.
- Concerns regarding equivalence with paper-based assessments.
- Time taken for students to familiarise with the computer interface.

If we return to the framework of the 10 principles at the beginning of this section, the themes identified in the research on formative assessment focus very strongly on the use of CBA to uphold principle number 9 – opportunities for self-assessment. This is well evidenced in the available research and is exemplified by the findings in Bull *et al*. (2006) and McGuire (2005).

However, this leaves any claim that CBA can enhance a number of the principles for formative assessment un-evidenced from the research. Only one study demonstrated the use of CBA in focusing how students learn (Peat and Franklin, 2002). Similarly, there was only limited evidence on how CBA helped place AfL as central to classroom practice (McGuire, 2005). Little research evidence is available on the use of CBA to support the remaining principles.

In examining these studies related to the formative use of CBA, the literature indicates that only one key aspect of formative assessment is significantly evidenced, namely the capacity of CBA for instant feedback and providing opportunities for self-assessment or reflection. It is clear that there is much less evidence available for how CBA supports the remaining principles outlined in this section.

Evidence for CBA improving the validity of assessments

The American Psychological Association (APA) 'Guidelines for Computer based Tests and Interpretations' state that "the validation of computer-based tests and protocols does not differ in kind from the validation of tests generally" (APA, 1986, p.19).

It is worth emphasising the importance given to validity in evaluating new forms of assessment:

> *The arguments, pro and con, regarding traditional and alternative forms of assessment need to give primacy to evolving conceptions of validity if, in the long run, they are to contribute to the fundamental purpose of measurement – the improvement of instruction and learning.*
>
> (Linn *et al*., 1991, p.20)

Russell, Goldberg and O'Connor (2003) provide a useful summary of some aspects of validity research since 1986 which cites evidence on the following areas:

- The inability to review or revise responses (this, in particular, is a feature of computer adaptive testing) has a negative effect on examinee performance.
- Graphical display issues, such a screen size and resolution, affect examinee performance.
- Familiarity with computers plays a role in test performance.

These three areas refer to Messick's (1989) concept of 'construct irrelevant variance' which becomes a recurring theme in the reviewed literature. These points also serve as a useful illustration of three very different sources of construct irrelevant variance:

- The method by which assessment items are sequenced.
- Aspects of screen display.
- The characteristics of the student in relation to ICT.

Sources of construct irrelevant variance in an on-screen assessment of ICT skills are also explored by Threlfall, Nelson and Walker (2007), who approach the analysis by examining 'sources of difficulty' and then identifying those that are linked to the construct, and those that are irrelevant to the construct.

One recurrent feature of CBA research are studies designed to yield comparisons with 'equivalent' paper-based tests – this is evaluating the traditional dimension of criterion-related concurrent validity (often referred to as cross-modal validity in the literature). A meta-analysis of such studies by Bunderson *et al*. (1989) demonstrated better performance in computerised tests in 3 cases; no difference in 11 cases; and better performance in paper tests in 9 cases. The meta-analysis revealed some potential reasons for the modal differences:

- Aspects of item delivery.
- Aspects of item presentation.
- The students' background characteristics – particularly in relation to ICT.

Note that these points concord with the findings by Russell *et al*. (2003) above. Some research has focussed on the students' characteristics:

> *In summary, establishing a model that fully accounts for test performance differences may be some time away, however it seems*

critical at this time to further this line of research. Based on our review and these results, we anticipate that computer familiarity is the most fundamental key factor in the test mode effect.

(Clariana and Wallace, 2002, p.601)

Huff and Sireci, (2001) examine issues regarding validity in computer-based testing. First they look at the evidence used in favour of CBA to enhance validity and conclude that most of these arguments centre on: 1) increasing construct representation, and 2) improving measurement precision.

They go on to state that the claims that computer-based testing can enhance validity can be traced to at least four current developments:

- Innovative item formats.
- Computerised-adaptive testing technology.
- Cognitively principled CBT design.
- Automated scoring.

In the same article, perceived threats to validity are also explored, namely construct under-representation and the introduction of construct irrelevant variance in CBT tests.

Regarding 'construct under-representation', some argue that CBA can improve the construct validity of a test in the case of assessments of problem solving skills (Ridgway and McCusker, 2003) and students' cognitive strategies (Nunes, Nunes and Davis, 2003). Some attempts have been made to provide a more empirical demonstration of construct validity in assessments of students' cognitive strategies (Wirth and Klieme, 2003).

From the discussion above, it is clear that there are many areas of 'construct irrelevant variance' to explore. Some inroads have been made with regards to identifying student background factors such as familiarity with ICT that have effects, but there clearly remain many areas of research activity left to explore in this area, particularly as CBA continues to evolve.

It appears that CBA introduces new sources of 'construct irrelevant variance' that, unmitigated, may reduce the validity of assessments. On the plus side, there is now a growing body of evidence indicating that CBA can facilitate the assessment of new constructs such as problem solving and meta-cognition.

Evidence for CBA facilitating access to educational assessments

The literature in this area can be divided into two further categories:

- The use of CBA to facilitate accessibility to assessments for individuals with disabilities.
- The use of CBA to facilitate access to assessments on-demand.

Taking the former, it is suggested that the increasing use of computer-based aids for those with disabilities make CBA an easier form of assessment to take advantage of these. For example, Bennett (1999) states:

From the perspective of task comparability, CBT offers substantial promise. One reason is that computers have become life-style accommodations for people with disabilities…an industry has evolved that produces dozens of alternative devices for getting information into and out of a personal computer.

(Bennett, 1999, p.181)

Empirical research is not evident in this area, which may be down to the small numbers of students with disabilities taking part in large-scale computer-based assessments. There is a warning associated with the use of features to enable accessibility, that un-checked they could evolve into threats to validity, that is, providing unfair assistance to particular students (Hansen, Mislevy, Steinberg, Lee and Forer, 2005) .

The latter aspect of accessibility, that CBA facilitates on-demand testing, can be illustrated using case studies. For example, an evaluation of CBA in undergraduate level Chemistry (Lowry, 2005) collected some qualitative feedback from students who indicated one benefit of CBA was the usefulness of being able to access the material at any time. The growth of the internet has clearly offered opportunities of online testing on demand:

Online students are able to take advantage of the accessibility of online assessment tasks from a variety of locations. They may receive valuable 'just in time' feedback from their teachers in order to make meaningful, timely decisions and judgements about their own learning.

(Northcote, 2002, p.623)

It would appear that there are further areas of empirical work to be done with regard to CBA improving accessibility to assessments for those with disabilities, particularly as increasing numbers participate in CBA programmes.

It would appear that commentators believe CBA has clear benefits in offering accessibility to those with disabilities. However, the lack of empirical research means that evidence is still awaited.

Evidence for CBA effects on student motivation

Harlen and Deakin Crick (2003) provide a useful framework for examining motivation through their meta-analysis of 19 studies linking motivation and testing. However, none of the studies examined involved the application of CBA. Even if it were assumed that findings would transfer to the CBA environment, they largely imply a negative association between testing and motivation. However, the hypotheses put forward by, for example, Bull and McKenna, (2003) and McKenna (2001) were that CBA improves student motivation. It has been difficult to find much evidence of motivational effects specifically associated with CBA, this is clearly an area that is ripe for further research.

Conclusions

The literature review has identified 5 themes associated with the evaluation of CBA.

In none of these areas was there comprehensive empirical evidence in the existing literature to back up the claims made for CBA.

Much of the evidence has emerged from case-study methodologies (particularly in the area of formative assessment), meaning that opportunities for generalisation are limited.

The two strongest themes are those of validity, which has been considered in a number of contexts, and formative assessment, which has evidence compiled from a number of case studies.

When the evidence gathered regarding the use of CBA in all five areas is scrutinised, a number of areas for further exploration and research activity emerge:

- Evidence primarily emerges for Higher Education contexts, suggesting that more work could be done to identify issues

specifically related to secondary or primary levels of education.

- There is a possible need to evaluate cost effectiveness in a more conclusive manner.

- It would be more comprehensive to evaluate CBA against the additional criteria for formative assessment from the Assessment Reform Group that were discussed in the evaluation of evidence for the formative use of CBA.

- Aspects of validity relating to construct-irrelevant variance could be explored in the context of computer-based tests.

- There is a need to provide more empirical evidence on the impact that CBA has in supporting access.

- Work could be undertaken to identify if there is a link between the use of CBA and student motivation.

- A critical and structured review of the 'grey-literature' about CBA.

It is also possible to derive a number of practical aspects from the literature reviewed here, which will be of use to those considering how CBA may or may not improve the assessment experience at their learning institution:

- CBA does not have a strong empirical basis for efficiency claims – therefore the literature would suggest caution if the prime motivation for the introduction of CBA is efficiency.

- CBA has a strong case for improving self-assessment opportunities (particularly in the case of Higher Education).

- CBA has a limited evidence base for facilitating full formative assessment practice – therefore the introduction of CBA alone is unlikely to lead to full scale adoption of formative assessment.

- The literature indicates that CBA does have the opportunity to facilitate access to educational assessments.

- There is weak evidence for the motivational effects of CBA.

These practical pointers, although primarily of interest to those currently considering the use of CBA in education, will be important considerations for researchers undertaking empirical work about CBA in the future.

References

APA (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC.: American Psychological Association.

Ashton, H. & Wood, C. (2006). Use of Online Assessment to Enhance Teaching and Learning: the PASS-IT Project. *European Educational Research Journal*, **5**, 2, 9.

Assessment Reform Group (2002). *Assessment for learning: 10 principles*. Assessment Reform Group. Retrieved February 26, 2009, from http://www.assessment-reform-group.org/CIE3.PDF.

Bennett, R.E. (1999). Computer-based testing for examinees with disabilities: On the road to generalized accommodation. *Assessment in higher education: Issues of access, quality, student development, and public policy*, 181–191.

Black, P. & Wiliam, D. (1998). *Inside the Black Box: Raising Standards Through Classroom Assessment*. School of Education, King's College London.

Brown, S., Race, P. & Bull, J. (1999). *Computer-assisted Assessment in Higher Education*. London: Kogan Page.

Bull, J. & McKenna, C. (2003). *Blueprint for Computer-assisted Assessment*. 1st ed. London: Routledge.

Bull, S., Quigley, S. & Mabbott, A. (2006). Computer-based formative assessment to promote reflection and learner autonomy. *Engineering Education*, **1**, 1, 8–18.

Burkhardt, H. & Pead, D. (2003). Computer-based assessment: a platform for better tests? *Whither Assessment*, 133–148. London: Qualifications and Curriculum Authority.

Bunderson, C.V., Inouye, D.K. & Olsen, J.B. (1989) The four generations of computerized educational measurement. In: R.L. Linn (Ed.). *Educational measurement*. 3rd ed., 13–103. New York: American Council on Education.

Clariana, R. & Wallace, P. (2002). Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology*, **33**, 5, 593–602.

Clariana, R., Ross, S. & Morrison, G. (1991). The effects of different feedback strategies using computer-administered multiple-choice questions as instruction. *Educational Technology Research and Development*, **39**, 2, 5–17.

Clarke, M.M., Madaus, G.F., Horn, C.L. & Ramos, M.A. (2000). Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies*, **32**, 159–181.

Conole, G. & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J Research in Learning Technology*, **13**, 1, 17–31.

DCSF. (2008). *The Assessment for Learning Strategy*. Retrieved February 25, 2009, from http://publications.teachernet.gov.uk/default.aspx?PageFunction= productdetails&PageMode=publications&ProductId=DCSF-00341-2008.

Ely, D.P. (1999). Conditions that facilitate the implementation of educational technology innovations. *Educational Technology*, **39**, 6, 23–27.

Fuchs, L.S. & Fuchs, D. (1986). Effects of systematic formative evaluation: a meta-analysis. *Exceptional Children*, **53**, 3, 199–208.

Gipps, C.V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: Routledge.

Hansen, E.G., Mislevy, R.J., Steinberg, L.S., Lee, M.J. & Forer, D.C. (2005). Accessibility of tests for individuals with disabilities within a validity framework. *System*, **33**, 1, 107–133.

Harlen, W. & Deakin Crick R. (2003). Testing and Motivation for Learning. *Assessment in Education: Principles, Policy and Practice*, **10**, 2, 169–207.

Huff, K.L. & Sireci, S.G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*, **20**, 3, 16–25.

Hunt, N., Hughes, J. & Rowe, G. (2002). Formative Automated Computer Testing (FACT). *British Journal of Educational Technology*, **33**, 5, 525–535.

Jean, S., Delozanne, E., Jacoboni, P. & Grugeon, B. (1998). Cognitive Profiles in Elementary Algebra: the PÉPITE Test Interface. *Education and Information Technologies*, **3**, 3, 291–305.

Johnson, M. & Green, S. (2004). *On-line assessment: the impact of mode on student performance*. Paper presented at the British Educational Research Association Annual Conference, Manchester, September 2004.

Linn, R.L., Baker, E.L. & Dunbar, S.B. (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*, **20**, 8, 15–21.

Loewenberger P. & Bull J. (2003). Cost-effectiveness analysis of computer-based assessment. *Alt-J – Association for Learning Technology Journal*, **11**, 2, 23–45.

Lowry, R. (2005). Computer aided self assessment–an effective tool. *Chemistry Education Research and Practice*, **6**, 4, 198–203.

McGuire, L. (2005). Assessment using new technology. *Innovations in Education & Teaching International*, **42**, 3, 265–276.

McKenna, C. (2001). Introducing computers into the assessment process: what is the impact upon academic practice? *Higher Education Close Up Conference*, **2**.

Messick, S. (1989). Validity. In: R.L. Linn (Ed.) *Educational measurement*. 3rd ed., 13–103. New York: American Council on Education.

Nicol, D. (2007). Laying a foundation for lifelong learning: Case studies of e-assessment in large 1st-year classes. *British Journal of Educational Technology*, **38**, 4, 668–678.

Northcote, M. (2002). Online assessment: foe or fix? *British Journal of Educational Technology*, **33**, 5, 623–625.

Nunes, C.A.A., Nunes, M.M.R. & Davis, C. (2003). Assessing the inaccessible: metacognition and attitudes. *Assessment in Education: Principles, Policy and Practice*, **10**, 3, 375–388.

Peat, M. & Franklin, S. (2002). Supporting student learning: the use of computer-based formative assessment modules. *British Journal of Educational Technology*, **33**, 5, 515–523.

Ridgway, J. & McCusker, S. (2003). Using computers to assess new educational goals. *Assessment in Education*, **10**, 3, 309–328.

Russell, M., Goldberg, A. & O'Connor, K. (2003). Computer-based testing and validity: A look back into the future. *Assessment in Education: Principles, Policy & Practice*, **10**, 3, 279–293.

Thelwall, M. (2000). Computer-based assessment: A versatile educational tool. *Computers and Education*, **34**, 1, 37–49.

Threlfall, J., Nelson, N. & Walker, A. (2007). *Report to QCA on an investigation of the construct relevance of sources of difficulty in the Key Stage 3 ICT tests*. Retrieved February 26, 2009, from http://www.naa.org.uk/libraryAssets/media/Leeds_University_research_report.pdf.

Topping, K.J., Samuels, J. & Paul, T. (2007). Computerized assessment of independent reading: Effects of implementation quality on achievement gain. *School Effectiveness and School Improvement*, **18**, 2, 191–208.

Welch, A.R. (1998). The cult of efficiency in education: comparative reflections on the reality and the rhetoric. *Comparative Education*, **34**, 2, 157–175.

Wirth, J. & Klieme, E. (2003). Computer-based assessment of problem solving competence. *Assessment in Education: Principles, Policy and Practice*, **10**, 3, 329–345.

Is CRAS a suitable tool for comparing specification demands from vocational qualifications?

Jackie Greatorex and Nicky Rushton Research Division

Introduction

Historically, unitary awarding bodies and the national regulator[1] monitored standards of qualifications between awarding bodies, over time and between cognate qualifications at the same level, and this work continues. A key reason for conducting such work is to avoid inequalities and inequities which would be created by the existence of easier routes to access further study or jobs.

Ideally standards are compared in terms of candidates' performance and in terms of the demands of the qualifications. When comparing new qualifications there is sometimes a lack of performance evidence[2] or assessment tasks[3] to form a robust sample from which generalisable research results can be drawn. In such cases comparability studies could focus on specifications[4] and the associated demands. However, studies restricted to one aspect of comparability (whether it be performance or demands) are limited.

One approach to comparing demands of qualifications is for experts to rate them on a scale of cognitive demands known as CRAS. CRAS was developed using academic qualifications. An issue deriving from its provenance may be that CRAS is not suitable for use with vocational qualifications which are different in nature and purpose to academic qualifications. Generally there are far more comparability studies about academic qualifications than VQ/VRQs[5]. In the present study we investigate whether CRAS is suitable for use in comparability studies which include VQs/VRQs.

Demands and difficulty

There is sometimes a lack of clarity about definitions of demands and difficulty.

In this article:

Task demands refer to the actions (usually cognitive) a task is intended to require of typical members of the target group of learners. For example, candidates might be required to recall familiar information. Task demands generally relate to individual summative assessment tasks such as examination items. But task demands could also be related to an individual classroom activity or similar.

Specification demands refer to the actions the specification is intended to require of typical members of the target group of learners in four areas: cognitive, affective, psychomotor and interpersonal. These specification demands might be explicit in the specification or they might be an underpinning ethos. For example, candidates might be required to recall information about a topic, empathise with another person's understanding of the topic, evaluate the other person's understanding to know what extra information they need and explain the relevant information to the

1. Currently the national regulator of the awarding bodies is Ofqual.

2. *Performance evidence* refers to students' work in the form of essays, artefacts, paintings, multiple choice responses and so on.

3. *Assessment tasks* refers to examination questions, assignments, briefs for work-based projects and so on.

4. The specification is: *The complete description – including optional and mandatory aspects – of the content, assessment arrangements and performance requirements for a qualification. A subject specification forms the basis of a course leading to an award or certificate. Formerly known as a 'syllabus'.* QCDA (undated)

5. VQ refers to vocational qualifications and VRQ to vocationally related qualifications. These are very broad categories. Many vocational qualifications in England are NVQs (National Vocational Qualifications which: *are designed to recognise a candidate's competence in the workplace. They provide a statement to employers of skill, competence and knowledge in a particular sector.* (OCR, 2009a).
Vocationally related qualifications generally focus on an occupation or occupational sector: *Vocationally-Related Certificates enhance knowledge and build upon candidates' skills in preparation for a job.* (OCR, 2009b).