



PERGAMON

Computers & Education 34 (2000) 37–49

---

---

**COMPUTERS &  
EDUCATION**

---

---

www.elsevier.com/locate/compedu

# Computer-based assessment: a versatile educational tool

M. Thelwall

*School of Computing, Wolverhampton University, Wolverhampton, WV1 1SB, UK*

Received 16 June 1998; accepted 8 November 1999

---

## Abstract

There are many types of Computer-Based Assessment in use today in higher education, from formative self-tests to summative final exams. CBA is used in various ways as an integral part of many currently used Computer Aided Learning environments. This paper surveys some of the reasons for using computers for assessment and some of the types of test in use, including evidence that coverage can extend beyond first year assessments, where it seems to be stuck in some disciplines. One project and one type are examined in more detail: randomly generated open access tests. With this type of test students are allowed to practise in their own time before sitting the same test for a grade. Evidence from one particular test, in statistics, was taken from computer logs, questionnaires and individual interviews and is used to show that the test style motivated students to study, and, for some students, directed their revision even when they were away from the computer. The results show that random-based tests can have a number of major advantages over fixed assessments, including: increased lifespan, security and flexibility, improved student motivation for study; and use as a learning resource. They also indicate that CBA does not have to be viewed in isolation from the learning environment in which it is situated, but can have an impact upon a student's study strategy, for example through increased revision. Taken together with previous studies and the survey of uses, there is now a body of evidence to suggest that CBA is an extremely versatile educational tool. © 2000 Elsevier Science Ltd. All rights reserved.

*Keywords:* Authoring tools and methods; Teaching/learning strategies

---

## 1. Introduction

Computer-Based Assessment has become common in UK universities. In the early days, CBA was restricted to text-based simple question and answer sessions written by programmers.

---

*E-mail address:* cm1993@wlv.ac.uk (M. Thelwall).

0360-1315/00/\$ - see front matter © 2000 Elsevier Science Ltd. All rights reserved.

PII: S0360-1315(99)00037-8

Later, authoring systems were developed to create more user-friendly products and wider access for academics (Whiting, 1985). Today there are free, easy to use products for creating online assessments (Derbyshire, 1999) and commercial products such as Question Mark Designer for creating flexible networked tests. Questions do not now have to be text-based but can also incorporate graphics and multimedia. Responses assessed can include multiple choice selection, graphical hotspot clicking, text, numerical and mathematical answers. CBA is now used for exams, mid-unit tests and diagnostic assessments. It is also found as an integral part of many Computer Assisted Learning packages. Perhaps now, with the creation of a national Computer Assisted Assessment centre in the UK, it has come of age.

One strand running through a number of computer-based assessment projects has been the ability of the software to randomly generate questions for the students or to randomly select a coherent subset of questions from a large question bank. This has been achieved in a variety of different ways, particularly in maths and the sciences, and has served different purposes. Random-based tests take more time to create but, once created, do have a number of practical advantages stemming from the easing of security concerns, and can have a much longer life. In formative assessments random tests can also provide large sets of questions for students to practise, with instant feedback. This paper looks at the types of CBA in use and the reasons for implementing them before focussing on randomised assessment and one project in particular.

## **2. Types of CBA**

Computerised assessment is used in many different contexts, and to perform different functions. It is often split into formative and summative assessment, but it is convenient here to subdivide it further as in Table 1.

Each of the test types in the table has a unique set of issues and constraints. The CAL quiz and open access test are the only types that do not have a natural written predecessor. Many CAL programs now have assessment as an integrated part of the package for the learner. For example the virtual campus Internet course has a set of three multiple-choice questions after every three or four slides of new material ([www.virtual.scit.ac.uk](http://www.virtual.scit.ac.uk)). A traditional predecessor of this is a lecturer asking questions of the class at various times. A common predecessor of the open access test is the availability of a number of past papers for practice, and a tutor willing to mark them.

It should be noted that there are two different ways in which assessment can be formative: by providing feedback on performance and by the skills gained performing the task. All of the above would contain parts of both to some extent, but in varying degrees. Diagnostic tests and grading tests would be primarily relying on feedback whereas exercises would aim to improve skills but the other formative assessments could contain a significant mix of both.

The assessments that are not purely formative have security implications, which due to the technology are different to, and more complex than, paper tests (Zakrzewski & Bull, 1998). Written tests can be stored in a sealed envelope, but electronic ones usually need to be stored on a network in advance of the test and therefore need to be protected from student access, yet also need to be tested before use.

Table 1  
A taxonomy of applications of CBA

Area	Type	Description
Summative	Exam	An assessment solely for grading purposes such as an exam at the end of a unit of study (Callear & King, 1997; Zakrzewski & Bull, 1998).
Formative/ summative	Grading test	An assessment for grading but which also provides feedback intended to direct future studies. Includes a mid-unit small test, or weekly problem sets (Callear & King, 1997; Zakrzewski & Bull, 1998).
Formative	Open access test	A grading test that doubles as a set of exercises because students are allowed to practice before sitting the test (Thelwall, 1998).
Formative	Self-test	An assessment designed to give feedback to a student on their progress with a section of a unit of study (OR test, <a href="http://www.cba.scit.wlv.ac.uk">http://www.cba.scit.wlv.ac.uk</a> ; Zakrzewski & Bull, 1998).
Formative	Exercises	A problem set designed to consolidate learning on a section of a unit of study (Whiting, 1985; Thoennessen & Harrison, 1996).
Formative	Programmed learning tool	A linear CAL package-based upon a question and answer session, as pioneered by Skinner (1968), ( <a href="http://www.scit.wlv.ac.uk/~cm1993/stats/">http://www.scit.wlv.ac.uk/~cm1993/stats/</a> ).
Formative	CAL quiz	A marked exercise integrated into a CAL package, for example a multiple-choice question presented after a slide containing new information (Beilby, 1993; Kelly, Maunder & Cheng, 1996).
Formative	Adaptive CAL quiz	A marked exercise integrated into a CAL package used to test the student but also used to adapt the teaching of the package to student weaknesses (Laurillard, 1993).
Formative	Diagnostic test	An assessment of prior learning taken before a unit of study (Appleby et al., 1997).

### 3. Computerisation objectives

Just as there are many types of computerised tests, there are also different educational motives for the introduction of CBA into a unit of instruction. There are also non-educational motives: reasons separate from concerns about the impact on the learning experience. These concerns are summarised as follows, with educational objectives as the starting point.

- Virtue — The computerised test would improve the learning experience for students.
- Necessity — A written test is not an option, for example in self-study CAL.
- Expediency — The same educational objectives would be met by both but other concerns promote computerisation
  - Efficiency — Automatic marking can save staff time.
  - Strategic — Testing is expected to have a positive impact on other areas such as increasing the take-up of CAL.
  - Policy — Institutional policy promotes CBA use.
  - Research/experimentation — The impact of CAL is evaluated.

### 3.1. *Virtue*

The most obvious way in which computerisation of assessment can be a virtue is through instant marking and feedback. It is believed that instant feedback is often more educationally effective than when delivered after a delay, possibly of days or weeks, for human marking (Dempsey, Driscoll & Swindell, 1993). In addition to this, quick feedback is clearly desirable for much formative assessment, particularly for diagnostic quizzes and self-tests when the results may be used to direct remedial studies. Open access tests can also promote understanding, particularly by weaker students, by encouraging revision. The student experience can also be indirectly improved as a result of the lecturer's easy access to a breakdown of test marks.

There have been concerns expressed about the effectiveness and scope of CBA in both formative and summative roles. Laurillard (1993) in the context of a simple adaptive CAL quiz raises the concern that it would not be able to do much to help develop conceptual understanding. Jones (1990) also doubts that CAL packages can build higher thinking skills if they are linear. One response to this has been constructivism and the creation of environments for exploration and learning (Grabinger & Dunlap, 1993). There is, however some evidence that formative computer-based assessments can aid retention of concepts (Bocij & Greasley, 1999) and subject mastery in some sense (Whiting, 1985).

One major concern raised about summative CBA is whether it can assess higher intellectual skills. It is in common use for first year students, but is less common elsewhere, with notable exceptions, such as medicine, where it is used in the final year of degrees and for the continuing professional qualifications of practising physicians. There are some skills, such as the ability to construct a logical argument in essay form, that CBA cannot assess, although there have been partially successful attempts to devise metrics to automatically mark essays (Christie, 1998). It is inconceivable to use a computer to assess large, important, predominantly creative projects such as a final year creative writing essay or a final year arts project. Computers are, however, used to assess some large creations, such as computer programs (Foubister, Michaelson & Tomes, 1997; Foxley, Higgins & Tsinitisifas, 1998), although a computer program clearly exists in a tightly defined environment and has a relatively small artistic component.

Many CBA tests are mainly or exclusively multiple-choice. But multiple-choice questions do not have to test facts, but well-written questions can require interpretation and probe understanding (Carneson, Delpierre & Masters, 1998), although not directly the ability to construct something. Here are some illustrative examples of ideas for multiple-choice type question stems that would not just test knowledge.

Given an unknown picture:

- Select two painters that were major influences on the artist...
- Pick a word that best describes the mood of the picture...
- Which of the following aspects of painting do you think the artist was most concerned about...

Given a section of prose:

- Select all the metaphors.
- Select all the similes.
- Identify the inappropriate metaphor. It is also easy to devise complex logic problems with multiple-choice answers. A simple example of this would be:

Which of the following is the correct negation of the statement: ‘All students in the class with brown shoes are female...’

Research has also shown that Multiple-choice tests can be as reliable as human marked essays (Kniveton, 1996). It is believed, therefore, that multiple-choice tests and computer-based assessments can have a place in the highest levels of education, but should be used in conjunction with other assessments to cover areas beyond their scope.

### 3.2. Necessity

In a CAL package designed with numerous short quizzes, it would be impractical and slow to manually mark the papers and therefore computerised marking or at least computerised presentation of the answer is the only option.

There are also other situations where it would be impractical to mark an assessment without computers and therefore these would not have been set previously. An example of this is the Introductory Maths for Business module at Wolverhampton University which, with over six hundred students, did not have a paper-based diagnostic test but adopted a computerised one.

### 3.3. Expediency

The impact of the method of assessment delivery upon students is not always the critical factor in the decision to adopt CBA. Paper and computer methods may be seen as equivalent in some cases, when other concerns dictate which is adopted.

Computers have been used for reasons of efficiency in marking multiple choice tests with OMR devices for many years, in school ‘A’ levels and a number of HE institutions continue to use the technique for large cohorts (Griffiths, Hancock & Yoffe, 1998; Stephens, 1994). Setting up and administering CBA is more time consuming than simply putting the questions on paper for a traditional test, but logically there must be a break-even class size beyond which marking time saved exceeds set up time, unless administration time per student is greater than marking time. If a test is re-used then this break-even point could be achieved over a number of years. CBA is therefore attractive for saving marking time with large classes. It may also serve to shift the exam burden from Lecturers to others, such as computer technicians administering the system (Zakrzewski & Bull, 1998).

Some CBA is promoted for strategic reasons. One such is its potential to save marking time which could be used to encourage lecturers to adopt it, opening the door to CAL (Bull, 1994). It may also be institutional policy to use CBA widely for certain exams, such as in Luton University. Related to this is its use by individuals or groups to evaluate various aspects of its impact upon the students and the university (Stephens, Bull & Wade, 1998).

#### 4. The need for randomisation

Under certain circumstances it is desirable to have a larger number of questions than would normally be needed for a single test, for educational reasons normally in a formative test or for security reasons in a summative test. A large question bank could facilitate exam preparation by allowing a lecturer to create a test by selecting questions, or the process could even be automated to create a randomly selected test. In formative tests, large question banks may mean that students are allowed to practice the tests as much as they like, unless it is undesirable in the case of a diagnostic test for example. For a summative test or exam, there are a number of possible security-related benefits.

- The risk of cheating is reduced or eliminated because every student has a different test.
- The same test can be used every year and for resits.
- All students do not have to take the test at the same time and there is no advantage in taking it first or last.
- Students can take the test at home or in their own time if they are trusted to do so.

The TAL project at Bristol University (Williams, Maher & Barry, 1999) is an example of an environment providing randomised access to large banks of questions. This is a web-based environment containing a large bank of multiple-choice questions to be shared among academics. Tests can be constructed by selecting topic areas that need to be covered. In order to get round the problem of sharing questions between different ability groups of students, there is a facility to set the test difficulty, which is automatically calculated on the basis of the performance of previous students.

Devising large banks of questions is time consuming if co-operation between educators cannot be achieved. For the multiple-choice questions that are commonly used in CBA, constructing effective distracters for a large set of questions can be a lengthy process (Carneson et al., 1998).

In some numerate subjects it is common practice to make the individual questions randomly generated questions. For example the two main CAL tools in Mathematics, MathWise (Beilby, 1993) and Transmath (Kelly et al., 1996) both include exercises as an important and integral part. The questions contained are often randomly generated for an effectively infinite problem set to allow students to practice as much as they like. Randomised tests of this kind have also been used in summative tests or exams in Mathematics (Davidson & Goldfinch, 1998; Foster & Crofts, 1997; Thelwall, 1998), Statistics (Simonite, Ells & Turner, 1998; Hunt, 1998), Computing (Kayssi, El-Hajj, El Assir & Sayyid, 1999), Engineering (Lloyd, Martin & McKaffery, 1996), and Physics (Thoennessen & Harrison, 1996). Automatically generated tests with random factors may, however, have a number of possible drawbacks over standard CBA.

- More time is needed initially to prepare a test with random factors.
- A programming language rather than an off the shelf CBA programme is normally needed although some CBA software includes provision for including random factors.
- There can be the loss of a sense of ownership for staff who do not programme or write the test themselves, but could have written a non-random test.

From an efficiency point of view, as long as there are large numbers on the module and the module does not change dramatically from year to year a test with random factors could save time compared to both written tests and CBA tests.

## 5. Constructing randomised test items

Various approaches are used to include random factors in the individual test questions. The Knowledge Space Project (ALEKS, 1997) uses the terminology of a test *item* which is a specific skill for which there are a number of questions that can be used to test the same. A similar approach can be adopted in many cases, constructing questions to test a particular skill or knowledge and then adapting them to produce a variety of questions that test the same thing. In some areas this is not appropriate and a more sampling approach needs to be taken. Samples of equally appropriate questions which test different things can be grouped together and a subset picked at random. As an example of this the Wolverhampton IT test includes questions on the functions of the buttons on the Word for Windows button bar. One question was written for each button and then the questions grouped into fours in terms of difficulty. One question out of each four would be picked at random each time the test is delivered. In this case slightly different knowledge items were being tested but the overall test difficulty would be approximately the same.

Number related tests can be the easiest to produce because random numbers can be generated and inserted into the question with the answer easily being calculated directly from the random numbers. The wording of the question and letters used can also be varied independent of the numbers which quickly produces hundreds of variations of questions testing the same skill. For example one of the Wolverhampton statistics test questions is on hypothesis tests. There are five different wordings of the question and these are descriptions of where the data comes from, for example fish lengths in cm or adult ages in years. There are also five basic variations of the data set with five different answers, but each variation of the data is further subject to a random factor with about 5 variations in such a way that the numbers change but the characteristics of the data don't. This gives 625 versions of the same question. Here is the text of two of them for comparison.

- Question 2, Version 1    2) A manufacturer of toughened glass claims that it can survive for 219 hours in the testing machine. A sample of 9 sheets of toughened glass were taken and it was found that the sample mean was 223.5 and the sample variance was 27. Set up and test null and alternative hypotheses to check the manufacturer's claim.
- Question 2, Version 2    2) A scientist claims that Bigwing butterflies have a 201 day lifespan. A sample of 16 butterflies were taken and it was found that the sample mean was 198.5 and the sample variance was 4. Set up and test null and alternative hypotheses to check the scientist's claim.

The students generally recognised in this case that the same skill was being tested by the

different variations of the question but that there was no short cut to learning the technique necessary to solve the problem.

There are areas in maths which lend themselves to randomly generated questions, such as algebra and calculus. One of the simplest questions in Transmath is a factorisation question such as “Factorise  $2a + 10$ ”. The numbers and letter here are randomly chosen and the answer is an algebraic expression:  $2(a + 5)$ . Transmath uses an algebra engine to check the correctness of the answers to the more complex questions.

It was discovered in the Wolverhampton project that tests could be designed by obtaining a number of past papers for the module and producing a prototype computer test-based upon these. This was then shown to the module leader for comment and amendments. This method was found to be more practical than involving the lecturer in the production of the material at an earlier stage as they were often uncertain about what was possible for their module or how to introduce the randomisation.

## **6. The Wolverhampton project**

The Wolverhampton University CBA project has been established for five years, and produces random-based tests. The original stated aim of the project was to promote the use of CAL by demonstrating the reliability and timesaving efficiency of CBA. It began with a single statistics test, the success of which led to the adaptation of the technology to the production of other tests and more attention being paid to the pedagogical implications of different designs. Four tests replaced written predecessors in maths, stats and IT, and two diagnostic tests for numeracy and programming were also created, one of which replaced a written version. The tests are used on six modules with about 1000 students in total each year. The tests have a common basic design, they are PC-based and run with over 80,000 variations, mark the questions only at the end of the exam and save the data to a central database.

Although the same technique and randomisation is used for the production of these tests they fall into three different groups: open access tests; a grading test; and diagnostic tests. The maths and statistics open access tests are available at all times for students to practise on, and count for assessment only when taken in the actual test session. The IT grading test is password protected and the students only use it for the actual assessment test. It was thought that the nature of the module meant that test practice would not benefit students because it would be more beneficial for them to spend the computer time using the package being learned. The other two tests are diagnostic tests, which are adaptive in the sense that they ask questions depending on the answers to previous questions. A few questions in one of the tests, the IT test, had no random factors. These questions required the user to demonstrate moving and resizing a window. In the context of a large test, where all the other questions did include randomisation, this was not considered a problem.

When all the questions in a test have random factors it is easy to get an improbably large number of the overall random factor for the whole test. For all the tests produced except the small computing diagnostic test the cumulative random factor exceeded the theoretical maximum of the method used to generate random numbers, which is just over 80,000.



For staff using the tests reliability and sufficient randomness to make the test still valid after several practises are important factors. Reliability of the program itself is necessary to avoid all the hassles of having to reschedule a test if the program crashes. The reliability of the marking is helped by the test methodology in the case of the open access tests. Students practising the test before the exams will discover mistakes in the test, if any, and these can be corrected before the final exam. Any that slip through the net can be caught by the lecturer from the feedback sheet and the marks manually adjusted.

## **7. Student perceptions and usage of the statistics test**

A study was made of one semester's cohort of statistics students and the computer log, a questionnaire and informal interviews were used to judge how it was being used. Although the overall aim of the project at the time of the creation of this test was to get CBA and CAL established, there were also specific educational objectives. One recognised problem with the module was that some students seemed to delay a serious attempt to begin their studies until it was too late, perhaps due to 'math anxiety'. Another common problem was of students from other disciplines becoming demoralised by being unable to understand the subject in the way that they could understand their 'own' subjects. The open access style, later enhanced by instant detailed feedback on each test, was an attempt to motivate both of these types of students in two ways. All students were expected to take the test for practice in the normal workshop time a couple of weeks before the real test and could practise in their own time thereafter. This first attempt was effectively a predicted grade, and served to disabuse some of the notion that they might scrape through without revising. The other motivating factor was that the fairly predictable nature of the test allowed students to set themselves targets to be able to answer certain questions correctly with a realistic expectation of the marks in the exam that counted. Some students as a result practised the test many times, increasing over time the number of questions that they could reliably answer correctly. The purpose of the computer log, questionnaires and informal interviews was to determine whether our impressions of the success of the methodology was correct or not.

A picture was built up of how the students were using the test using the combined data sources. The computer log recorded 1099 accesses to the programme and 168 actual tests. This gives an average of 5.5 test uses per student before the real test. The average figure was 3.3 practises per student from the 138 questionnaires collected and the difference could be explained by students "having a look" at the test without seriously practising it. Overall it is thought that the students probably practised the test on average at least four times on paper or on the computer before sitting it for real. One related result from the informal interviews was that some students were also taking home printouts to use for revision. This is particularly interesting because the test was having a wider impact upon how the students were studying the module, even away from the computer.

Interviews with students showed three distinct strategies for dealing with the test: not practising at all; practising on the computer; and practising on paper with printouts. Only a small minority did not practise at all. These seemed to be students who also did not turn up

for the lectures or workshops. Students who practised on the computer almost without exception made reasonable gains over time, as recorded by the log and only one student did worse on their test than in their first practise. “I hate computers and statistics [but] this is a compulsory module for Psychology” said one, “I practised six times, improving from a [25% Fail] to a [45% Pass] in the final test.” The log showed that another student, Melanie, took the test fourteen times, steadily improving from a C+ to an A+. Table 2 shows how her marks progressed over time.

Another student with a 75% score said that he had got 25% on his first try but had taken a number of printouts home and got up at 6 a.m. on the morning of the test to practise with them.

Student perceptions tied in with the log data, an overwhelming majority believing that the practice helped them. The questionnaires showed that 86% thought that the test helped them to learn against only 5% who thought that it did not with 9% undecided. Also 65% thought that it had motivated them to revise more against 14% who did not and 21% undecided, again showing a wider impact on studies.

The advantages of being able to practise and learn from the test programme were somewhat offset for some students by the partly correct perception that they would not get partial marks for partially correct answers, and that the computer was therefore a harsher marker than a lecturer. Despite this, at the end of the test 55% preferred computer-based assessment to traditional assessment against 12% who didn't and 34% undecided. Only 6% said that they did not trust the computer to mark their work at the end. In the history of the test, only one student has claimed that they failed the test only because it was computerised, and they also failed a written substitute test that they were allowed to take. As a result of this, and experiences with students with no previous IT experience, we have no concerns that the technology is a real obstacle to any student.

Although the grades on the test improved over the previous written version and exam grades have slightly improved, it is not possible to give a definitive statement of success because of the heterogeneous nature of the student base and the differing annual base skill level. Nevertheless, it is believed that the overall contribution is positive. After a discussion, a decision concerning whether to replace the final exam with another similar computerised test was resolved by keeping the written version because of the essential component of the course that could not be assessed by a computer: the ability to write a coherent and logically correct report on the results of a statistical calculation or test.

There are two factors that are believed to have contributed to the success of the statistics test, factors that should be taken into account when designing any similar test program: usability and feedback. Ease of use helped encourage students to use the test. Students found the statistics test easy to use (86% in the questionnaire) and also found the feedback useful

Table 2  
Melanie's marks on the test (out of 48)

Test	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Score	27	31	29	36	42	39	44	48	40	46	44	48	44	47

(91%). The original version of the test did not have feedback but student requests prompted its creation and it is a very popular feature. The users can print out the feedback and take it home to analyse their mistakes. The combination of ease of use and good feedback was an important factor in making the test a success with the students.

## 8. Conclusions

Despite the large number and variety of methods of computer-based assessment in use in higher education today, there are still misconceptions and unanswered questions concerning the role that it can play in the learning process and its possible scope for summative assessment. There is a range of literature covering different aspects of this, but the variety of types of assessments and disciplines in which they are used does not help to clarify the issue. It is easy to see that there are some things that a computer cannot assess and that tests of factual knowledge are not difficult to computerise but this can lead to a dismissing of the potential of CBA. There is, however, evidence that even the commonly used multiple choice questions can test higher intellectual skills and that in certain circumstances multiple-choice exams can be as reliable for assessment purposes as essays.

Different styles of assessment have different impacts upon students, and the randomly generated open access tests attempt to provide an inexhaustible supply of formative assessment in addition to their eventual use for summative assessment. The random element can give them a longer life-span, over five years now for the Wolverhampton statistics test, and allow them to double as a learning resource. Results of the survey into the open access statistics test showed that it is popular with students, it drives many to revise more and that it is used by a number to direct their studies, away from the computer. It therefore seems that it is able to provide one of the sometimes forgotten essential components of education: motivation. This is likely in this case to be some form of extrinsic or 'achievement' motivation, rather than a desire to understand the subject, but does point the way to conceptualising computer-based assessment not as separate from the rest of the education process but as situated in a learning environment and able to affect the way in which other materials are perceived and used. This, in conjunction with other evidence discussed and the taxonomy of uses, shows the versatility and potential richness of CBA for education.

There are a great many different ways of using computer-based assessment with students, from a diagnostic test given before studies begin, to formative assessments during study, mid-term tests and the final exam. Computerisation of assessment can enhance the value of education in certain circumstances, either because it is intrinsically better than paper assessment or because the assessment would be impractical without it. Computers cannot be used for all assessments because some educational outcomes, particularly those with a creative component, are too difficult to measure with the defined rule that a computer program needs. The point here is that CBA is *different* to and, in the information age, *complementary* to traditional assessment methods.

## References

- ALEKS. (1997). (<http://aleks.uci.edu/>).
- Appleby, J., Samuels, P., & Treasure-Jones, T. (1997). A knowledge-based diagnostic test of basic mathematical skills. *Computers and Education*, 28(2), 113–131.
- Beilby, M. (1993). Mathwise: A computer-based learning environment. (<http://www.bham.ac.uk/ctimath/reviews/nov93/mathwise.pdf>).
- Bocij, P., & Greasley, P. (1999). Can computer-based testing achieve quality and efficiency in assessment? *International Journal of Educational Technology*, 1(1) <http://www.outreach.uiuc.edu/ijet/>.
- Bull, J. (1994). Computer-based assessment: Some issues for consideration. *Active Learning*, 1, 18–21.
- Callear, D., & King, T. (1997). Using computer-based tests for information science. *Alt-J*, 5(1), 27–32.
- Carneson, J., Delpierre, G., & Masters, K. (1998). Designing and managing multiple-choice questions. <http://www.le.ac.uk/castle/resources/mcqman/mcqman01.html>.
- Christie, J. (1998). Computer-assisted assessment of essays. In *Proceedings of the second annual computer assisted assessment conference* (pp. 85–89).
- Davidson, K., & Goldfinch, K. (1998). How to Add — VALUE. <http://www.icbl.hw.ac.uk/ltidi/evalstudies/eshowtoadd.htm>.
- Dempsey, J. V., Driscoll, M. P., & Swindell, L. K. (1993). Text-based feedback. In J. V. Dempsey, & G. C. Sales, *Interactive instruction and feedback* (pp. 21–54). NJ: Educational Technology Publications.
- Derbyshire, H. (1999). The CASTLE (toolkit. <http://www.le.ac.uk/castle/>).
- Foster, M. G., & Crofts, A. (1997). Webtest and mathematics service courses. *CTI Maths and Stats*, 8(2).
- Foubister, S. P., Michaelson, G. J., & Tomes, N. (1997). Automatic assessment of elementary standard ML program using Ceilidh. *Journal of Computer Assisted Learning*, 13(2), 99–108.
- Foxley, E., Higgins, C., & Tsinitisifas, A. (1998). The Ceilidh system: a general overview. In *Proceedings of the second annual computer assisted assessment conference, June* (pp. 140–145).
- Grabinger, R. S., & Dunlap, J. C. (1993). Rich environments for active learning: a definition. *Alt-J*, 3(2), 5–34.
- Griffiths, P., Hancock, P., & Yoffe, O. (1998). Optical mark reading and the management of student records. In *Proceedings of the second annual computer assisted assessment conference* (pp. 39–47).
- Hunt, N. (1998). Computer aided assessment in statistics: the CAMPUS project. *Alt-J*, 6(2), 58–67.
- Jones, T. (1990). Towards a typology of educational uses of hypermedia. *Lecture Notes in Computer Science*, 438, 265–276.
- Kayssi, A., El-Hajj, A., El Assir, M., & Sayyid, R. (1999). Web-based tutoring and testing on a computer networks course. *Computer Applications in Engineering Education*, 7(1), 1–7.
- Kelly, A., Maunder, S., & Cheng, S. (1996). Does practise make perfect? Using TRANSMATH to assess mathematics coursework. *Active Learning*, 4, 30–36.
- Kniveton, B. H. (1996). A correlational analysis of multiple-choice and essay assessment measures. *Research in Education*, 56, 73–84.
- Laurillard, D. (1993). *Rethinking university teaching: a framework for the effective use of educational technology*. London: Routledge.
- Lloyd, D., Martin, J. G., & McKaffery, K. (1996). The introduction of computer-based testing on an engineering course. *Assessment and Evaluation in Higher Education*, 21(1), 83–90.
- Simonite, V., Ells, P., & Turner, W. (1998). Using IT to generate individualised coursework questions and solutions for an introductory course in statistics and probability. *CTI Maths and Stats*, 9(1), 16–18.
- Skinner, B. F. (1968). *The technology of teaching*. New York: Meredith.
- Stephens, D. (1994). Using computer assisted assessment: time saver or sophisticated distraction? *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.
- Thelwall, M. (1998). A unique style of computer assisted assessment. *Alt-J*, 6(2), 49–57.
- Thoennessen, M., & Harrison, M. J. (1996). Computer-assisted assignments in a large physics class. *Computers and Education*, 27(2).

- Whiting, J. (1985). The use of a computer tutorial as a replacement for human tuition in a mastery learning strategy. *Computers and Education*, 9(2), 101–109.
- Williams, J. S., Maher, J., & Barry, M. (1999). Introduction to the test and learn examination system. <http://www.tal.bris.ac.uk/Intro.htm>.
- Zakrzewski, S., & Bull, J. (1998). The mass implementation and evaluation of computer-based assessments. *Assessment and Evaluation in Higher Education*, 23(2), 141–152.