

Computer-assisted Assessment

Douglas Chalmers, Glasgow Caledonian University
and Dr W. D. M. McAusland, University of Aberdeen
Edited by John Houston and David Whigham, Glasgow Caledonian University

1	Introduction	2
1.1	What is computer-assisted assessment?	2
1.2	Why assess students?	2
1.3	What are the justifications for using CAA?	3
1.4	What are the disadvantages of using CAA?	4
2	Main content	5
2.1	Deciding whether you want to use CAA	5
2.2	What are the implications for different participating groups?	5
2.3	The desirability of question banks	7
2.4	What sort of questions can be used in CAA?	8
2.5	The opportunity for reusing questions	9
2.6	Plagiarism and cheating	9
2.7	Guessing	10
2.8	Using CAA to evaluate questions and areas of the module	10
2.9	Choosing your CAA system	10
2.10	Research and useful links	11
2.11	Examining available software	12
3	Case study: Macroeconomics at the University of Aberdeen (the 'Aberdeen case')	13
3.1	Background	13
3.2	The learning environment	14
3.3	Objectives and constraints	14
3.4	Practical implementation	15
3.5	Evaluating the experience	17
3.6	Final advice	18
4	Where next?	19
	References	19

1 Introduction

1.1 What is computer-assisted assessment?

Computer-assisted assessment (CAA) refers to the use of computers to assess students' progress. The assessments can vary in *format*: either consisting of a pre-printed paper test on to which students mark their responses, which are then processed automatically using an optical mark reader; or involving the direct input of students' responses into a computer terminal.

Computer-based assessments may be stand alone and specific to certain machines within a computer lab; based on a local network (intranet); or, as is increasingly common, web based.

The *nature* of the assessments may also differ. They can be *diagnostic* – to determine students' knowledge prior to starting a course, potentially allowing amendments to their specific course design.

They can be *formative* and include ongoing feedback either during the assessment or after.

They may be 'scored formative', allowing ongoing comparison of a student's progress over a period of time, possibly replacing an end-of-term summative assessment.

Alternatively, they may be *summative*, contributing to a student's end-of-year mark.

Depending on circumstances, such tests can be either supervised or non-supervised, with the option of allowing students to check their own progress through self-assessment. Although more commonly used for testing lower-order skills (such as knowledge, understanding and application), when properly formulated they can also be used for testing higher-order skills (analysis, synthesis and evaluation). Their nature allows the automation of what was previously a very time-consuming task: that is, marking and monitoring progress.

Like other tools, a properly constructed CAA package can be very efficient in achieving its designated tasks. The case study in section 3 (the Aberdeen case) exemplifies a procedure that was judged to be an overwhelming success. Unsurprisingly, if badly designed, or expected to fulfil inappropriate aims, a CAA framework will not achieve its potential.

1.2 Why assess students?

A number of reasons may be given to justify the use of assessments and need to be borne in mind when considering the appropriateness of CAA tests. They may include those shown in Box 1.

Box 1 Reasons for assessing students

- Helping the student learning process through formative assessments.
- Grading students' work through summative assessments.
- Being able to regularly evaluate student progress.
- Helping to monitor how effective current teaching and learning strategies may be.

It is also true that if topics do not have some sort of assessment procedure, students may not apply themselves to the topics with the preferred vigour.

1.3 What are the justifications for using CAA?

TOP TIPS

Before the adoption of any CAA system, ensure that full discussions take place to consider the aims of such adoption, then move on to a consideration of the practical options.

No one has as yet found the Holy Grail of assessment. In considering CAA as one of many possible methods, we need to weigh up the advantages and disadvantages for the different interest groups involved – lecturers, students, administrators and support staff. We should also consider whether any further move towards CAA is primarily pedagogically driven or simply resource driven. It is suggested in section 3 that the extent to which it will enhance student learning should be the paramount criterion, rather than how much time and other resources it will save. Given the differing forces that may be at work, it is therefore useful to consider the advantages and disadvantages of CAA from the perspective of pedagogy and administration.

Pedagogical advantages of CAA

If a successful system is designed and implemented, the claims of CAA are strong (see Box 2).

Box 2 Pedagogical advantages of CAA

- It enables the assessment of a wide range of topics very quickly, with an associated reduction in the time that lecturers dedicate to marking.
- The need for double marking is totally eliminated. This time and resource saving allows more regular assessment than might otherwise have been possible, consequently enabling more detailed knowledge of students' progress and quicker identification of problems.
- Tests can be tailored to match students' abilities and, with formative assessments, it may be possible for content to be varied automatically as the test itself progresses, matching students' weaknesses as they emerge during the test ('computer adaptive testing').

Sometimes, as in the Aberdeen case, CAA can fill a gap that neither lecturers nor smaller tutorial groups could otherwise fulfil, given resources. It is also argued that the adoption of the appropriate approach can lead to 'deep learning' in addition to more transient gains.

In terms of advantages to students, in addition to their being more closely monitored by academic staff, it is also possible for them to receive feedback during, or immediately after, a formative assessment. This allows students to self-assess and check their progress at their own

pace, hopefully turning assessments from events to be feared and endured, into ones that may even be enjoyed. As well as possibly changing students' perceptions of assessment, it may be the case that learners are more effective when they take note of externally provided feedback. Finally, in relation to the success and appropriateness of specific modules, feedback from students on the quality and pace of the course can be made an integral part of the assessment, allowing course evaluation by academic staff to be undertaken more easily. Again the Aberdeen case gives real-life instances of this.

TOP TIPS

Statistically evaluate assessments as soon as possible to spot problem areas of the course.

Administrative advantages of CAA

Again, a properly constructed and implemented CAA scheme can offer many advantages to institutions (Box 3).

Box 3 Administrative advantages of CAA

- The saving of time in supervision, invigilation and marking, and a reduction in subjectivity and human error in the marking process itself.
- When dealing with large groups of students, the time and resource saving can be of a significant order.
- Given the computer-based nature of the exercise, substantial reductions in printing costs should be achieved when assessments are updated or altered.
- Statistical evaluations of the results can be generated automatically, with no additional keying in of marks, thus reducing evaluation time.
- Successful integration into, and co-ordination with, student records and university information and management systems.

1.4 What are the disadvantages of using CAA?

Again it is useful to look at these from a pedagogical and an administrative point of view.

Disadvantages from a pedagogical point of view

These tend to relate to two issues: first, quality; and secondly, the initial set-up time and learning curve. In terms of quality, the construction of valuable objective tests (often the 'entry level' for a CAA) needs adequate staff training and will be initially time-consuming. An injudicious use of multiple-choice questions can lead to a focus on the testing of a superficial level of understanding, with consequent cynicism from staff and students alike. Although it is certainly possible to test higher-order levels of understanding, this is a more difficult task within CAA and again needs a high investment of initial time, together with the consequent higher level of quality assurance.

Objective tests cannot test a student's ability to communicate, or their propensity for original thinking, and thus should only be one of a battery of approaches adopted to test students' knowledge. Some issues may also exist regarding gender bias in multiple-choice questions. It has been claimed that females may be disadvantaged by such a format, although in a very useful review of literature on the subject, McKendree (2002) suggests that current evidence tends at most to suggest a weak systematic link.

A further disadvantage of CAA systems is that additional training in terms of IT and assessment design may be required for invigilators, given that the nature of possible issues which students may raise in a computer-based assessment are different from those which might be raised during a similar paper-based exercise.

In terms of student disadvantage, CAA tests also rely on a certain level of IT competence (which should never be overestimated, despite student familiarity with e-mail or web-based bulletin boards). If used as part of a distance learning approach, or otherwise relying on off-campus access, there may be additional issues of access to adequate computers or reliable and fast internet connections.

As with other computer-based approaches, issues may be raised for students with special needs. Finally, plagiarism may also be an issue unless adequate safeguards can be put into place regarding the authentication of students' identities. Additional measures may likewise have to be established to discourage and prevent cheating. These security measures would, of course, vary with the specific assessment format and could range from

TOP TIPS

Make sure that students are able to use the software package you have adopted well before it is used under exam conditions.

precautions such as loading the exam questions on to the server at the last moment in order to prevent unauthorised access, to additional encryption facilities being used to prevent unauthorised access to question and answer files.

Disadvantages from an administrative point of view

Similar to the higher set-up time in terms of academic preparation, a high level of co-ordination and co-operation will be needed between learning facilitators, often crossing several departmental boundaries. The administrative time in initially realising and implementing a scheme can be expected to be heavy.

TOP TIPS

If using CAA under exam conditions, always have some spare machines in the lab. If an individual machine crashes within 15 minutes, ask the student to restart on a spare machine. If after 15 minutes, allow them to complete a paper copy of the assessment.

Although breakdown of published software appears to be comparatively rare, the same cannot be said of individual computers within a computer lab, or indeed the local networks of which they are part. Likewise, the physical demands on IT staff, computer hardware and lab space if CAA schemes are to be conducted simultaneously on a large scale can be substantial.

All systems of CAA differ from each other, and in some cases may not integrate well with existing managed learning environments if already adopted. Others may have difficulty, for instance, in fully implementing anonymous marking – a standard approach is increasingly being adopted across institutions. Anonymous marking is very difficult to ensure, if CAA is invigilated.

2 Main content

2.1 Deciding whether you want to use CAA

In deciding whether to use CAA within a teaching environment, it is important to start by considering all the advantages and disadvantages listed above. Although some form of CAA is increasingly becoming the norm in most learning institutions, it is important that academic aims determine the framework of assessment that is adopted and not vice versa. In the Aberdeen case, by mastering the CAA programme, students were perceived to have developed a range of transferable skills of value in employment, together with more generic computing skills.

2.2 What are the implications for different participating groups?

Lecturers

Although details will obviously vary in each specific case, the responsibilities of lecturers and other academics with specific responsibility for CAA might typically consist of those shown in Box 4.

Box 4 Implications for lecturers

- Gaining access to the system by means of a password.
- Creating and editing questions.
- Choosing an appropriate range and style of questions for assessments.
- Setting usage restrictions on students.
- Subsequent to the assessment, monitoring the outcomes in terms of individual students, or groups of students, or the module as a whole.

In terms of training, CAA will demand a higher competence in IT issues than might otherwise have been the case, together with an understanding of pedagogical issues relating to the construction and composition of multiple-choice and objective test type questions, and their appropriateness in different circumstances. Integral to the advantage of CAA systems is the production of quality feedback – particularly when students make errors. This is perhaps one of the most time-intensive aspects for academics.

Students

As with learning facilitators, the implications for students will vary depending on the specifics of the system and the uses to which it is being put. However, typical usage might involve logging on to the system using a password (either on campus or, if permitted, from elsewhere), and navigating to the specific assessment using a web-type interface (such an approach allowing platform independence). Although this sounds very simple, problems that need to be addressed include basic ones such as whether students can use their log-on IDs in each location of a multi-venue assessment (for example, log-on IDs are sometimes restricted to some parts of a university network, while exams may be held in another).

If a formative assessment is being undertaken, students may have the option of working unsupervised at their own pace, receiving feedback as they progress and possibly retaking questions several times until an adequate understanding results in correct answers.

If the test is of a summative nature, this will probably be held in a supervised environment with a specified completion time, and final results either hidden from students or displayed at the end. As Whigham and Houston (1999) have shown, allowing students to monitor their own progress throughout an assessment will have implications for whether they engage in minimalist behaviour (for example, stopping once a threshold pass mark has been achieved).

In terms of training and ability, such assessment will demand a certain level of IT skills, together with an understanding of the specific format and demands of the CAA tests used by the department. In the Aberdeen case, for example, it specifically involved the construction of Excel worksheets, the successful completion of which required study in addition to that done in timetabled lab hours. This work was complemented by the students' use of web-delivered materials.

Students with special needs

As is the case with other computer-based work, visually impaired students or those with other special needs may need alternative versions of the given assessments, such as large print, spoken versions or Braille. Useful guidelines that deal with this issue have been produced by Hill (2000) and can currently be found at <http://www.disinhe.ac.uk/library/article.asp?id=33>.

IT staff

Although this may be considered as only a university-level resource matter, a higher use of IT will result in greater demands on IT staff. This was crucial in the Aberdeen case, where services of IT staff were ‘bought out’ to customise specifically the software package found most appropriate.

In general, a sympathetic approach will thus be needed from IT staff – particularly at ‘mission critical’ periods, such as scheduled online tests. Although ‘just-in-time production’ may be a possible approach to some aspects of higher education, such an approach would be very ill-advised in this area. Indeed, until full confidence is reached in the technology of a specific institution’s CAA system – the viability of the network being the crucial factor – it may be advisable to have a ‘Plan B’ in reserve. This might involve either rescheduling or possibly producing paper versions of proposed assessments in the case of unwanted problems such as computer failures.

Administrators

Again, it would be wrong to regard the effect of CAA on administrators simply as a resource matter. Successful use of CAA systems may mean a change in the responsibility of administrators regarding various aspects of the assessment process. While responsibility for pedagogical quality must lie with academic and quality assurance staff, it is administrative staff who should normally be expected to implement any day-to-day changes or developments requested. These might involve the electronic setting or design of assessments following the choice of material by lecturers. This will require training and an understanding of certain pedagogical issues regarding design.

TOP TIPS

Ensure that the division of responsibilities between different interest groups for updating and amending assessments is understood before using CAA live.

Invigilators will likewise need additional training in aspects of assessment design and IT issues in order to be able to differentiate between questions about the structure of an exam, which are valid, and questions about the exam’s contents, which it would be inappropriate for the invigilator to answer .

2.3 The desirability of question banks

The core of any robust system of CAA is normally the creation of appropriate, user-chosen question banks to be built upon over time. While this is undoubtedly the most time-intensive aspect of creating a viable CAA system, its success will be crucial to the success of the approach. It is no exaggeration to suggest that, as resources, the questions themselves are more important than the initial system, and will normally last longer than the system in which they are used.

The existence of such a bank of questions will also allow their reuse in appropriate circumstances, ensuring major time savings. Given the time-intensive nature of creating a question bank from scratch, you may wish to examine some of the reputable paper-based ones that are increasingly being produced by the major textbook publishers and see if these resources can be adapted or amended for your own use, subject to copyright considerations.

In judging the suitability of questions already available, it is important to consider the ultimate aims in terms of the level of skills to be assessed (lower order/higher order) and how/if this can be done with the material at hand. If not, can the material be amended in a satisfactory manner? Does the question material include satisfactory feedback? If not, can this be produced to fill the existing gap? How difficult would this be to do, given the specific package that is being used?

2.4 What sort of questions can be used in CAA?

Objective testing

Evidence to date would suggest (Zakrzewski, 2002) that objective testing is the most commonly used form of CAA. However, it would be wrong to underplay the variety of question types that can fall under this rubric. In addition to ‘traditional’ multiple-choice questions (MCQs) of the ‘choose one from the list’ type, it covers the types listed in Box 5.

Box 5 Alternative MCQ types

- True/false questions.
- Assertion-reason questions (combining elements of MCQ and true/false).
- Multiple-response questions (involving the selection of more than one answer from a list and picking from a list of options x, y, z, etc.).
- Text/numerical questions (involving the input of text or numbers at the keyboard).
- Ranking questions (requiring students to relate items in a column to one another, thus testing the knowledge of the order of events).
- Sequencing questions (requiring the student to position text in a given sequence, which is particularly good for testing methodology (McKenna and Bull 1999)).

Many sources can offer guidelines as to the construction of robust multiple-choice questions (Haladyna and Downing, 1989; Haladyna, 1994; Heard *et al.*, 1997).

Testing higher-order skills

The preponderance of objective testing does not mean, however, that CAA is only useful for testing lower-order skills. While objective testing does necessitate predetermined answers, these need not be limited to single-response options. Used imaginatively, more complex combinations comprising the appropriate answers to given questions can be constructed. Issues of design and creativity enter here, suggesting that time needs to be afforded to this issue, and a collective approach from a teaching team may allow the emergence of the best possible outcome. Details of websites where this is attempted are given in the References section.

Computer adaptive testing

One flexible form of objective testing useful in formative assignments is known as computer adaptive testing. Here, success in answering questions as they appear has an effect on the content and sequence of subsequent questions. In this way, the appearance of too many overtly easy or too many over-difficult questions can be prevented.

Methods other than objective testing

Assessment systems using methods other than objective testing are currently being developed by various institutions, although as yet few focus directly on economics. The Aberdeen case is one example. Other examples include Whigham and Houston (1999), and Dowsing and Long

TOP TIPS

Start constructing your question base well before you aim to get your CAA system up and running. Designate a team of people to work on this (ideally adherents of CAA).

Spend some time carefully cataloguing and indexing your question base – it will pay massive dividends later.

(1999) who have developed approaches to business data analysis using spreadsheets. These approaches are interesting in that they consider both the student's output and the 'event stream' – the order in which it was produced. Some areas of current research also focus on the question of assessing essays through CAA, interesting results being achieved in combining peer assessment with other forms of CAA (Davies, 2001).

Obviously the varying complexity of specific assessments will have different time implications in terms of students successfully completing their work.

Student response systems

Also worth noting is the interesting use of a 'student response system' – essentially a type of real-time interactive response that involves the students' use of infrared handsets (rather like TV remote controls) to respond to questions embedded within PowerPoint presentations. By use of these, the facilitator can request a response on the issues being currently examined and this real-time feedback is displayed via the data projector. The facilitator may thus assess the understanding of those in the audience before deciding whether or not to proceed (Nicholls, 1999).

2.5 The opportunity for reusing questions

By its nature, CAA allows the building of a resource bank of questions that can be reused. In fact, such flexibility should imply that decisions as to how questions are used are made after the questions have been entered into the question bank itself. Reuse can be appropriate in several ways. Under self-assessment, a student can practice by reusing the same material. Following its use in exams or summative assessments, material can be made available to students on an open access basis. Questions can also be 'pre-tested' in a formative assessment, allowing better subsequent construction of summative assessments. Similarly, the Aberdeen case gives examples of trailing the use of the 'Question Mark' software package before its summative use.

Ideally, questions should be capable of being easily categorised to allow cataloguing and subsequent choice by academic staff for use within assessments. With enough imagination, it is also possible that they can be reused in additional, different software formats where appropriate (although students may prioritise and practise old questions in order to pass their exam). Interesting work on the issue of reusing questions is currently being carried out at Heriot Watt University (Crofts 1999).

2.6 Plagiarism and cheating

Plagiarism is not a new problem, although it appears to be an increasing issue in terms of the easily accessible material capable of being downloaded from the web. CAA, under supervised conditions, should raise no qualitatively different issues regarding plagiarism from those normally encountered regarding student identity.

A number of steps can be taken to minimise cheating within summative assessments. The randomising of questions is one approach that can be adopted – either with questions appearing on different machines in different sequences, or with similarly graded but different questions being undertaken by different students.

Other practical measures can be adopted, such as customised screens being attached to computers to prevent those taking the exam from seeing other students' terminals. One study, testing the impact of regular formative assessments on groups of 40 students (up to 20 minutes per week), observed that 'time pressure, shuffling questions and unsynchronised starting times,

made it ineffective for students to look at others' screens' (Mulligan, 1999). Importantly, for the period of the assessment, the ability of students to use other computer-based facilities (such as wider internet access and e-mail) may have to be limited on the machines on which the assessment is being taken, in order to prevent students using these facilities to cheat.

2.7 Guessing

A fear about any type of objective testing, whether paper or electronically based, is that it will allow unwarranted high scores through intelligent or lucky guessing rather than knowledge of the subject. While the effect of guessing cannot be totally eliminated, a point worth making is that the effectiveness of students adopting this strategy diminishes in relation to the length of the assessment. Additionally, if wished, the use of corrective scoring may largely eliminate the effect of guessing. In corrective scoring, incorrect answers can be penalised by the adoption of the following scoring system: +1 for the correct answer, -1 for an incorrect answer and 0 for no answer.

The adoption of a corrective scoring system is not uncontroversial, however. One school of thought is that a well-constructed assessment, covering a sufficiently wide range of the course and with questions of varying difficulty, should minimise the distortions caused by guessing (McKendree, 2002). The disadvantages of negative scoring are that it takes students longer to complete an assessment, and that answers based on good partial knowledge (the development of which we may wish to encourage) may be discouraged. Whatever approach is adopted, it is vital that students are informed of the approach adopted towards marking, in order that they can adapt their strategies suitably.

2.8 Using CAA to evaluate questions and areas of the module

One of the uses of the automatically generated statistical analyses that can result from CAA tests may be to help academics identify and eliminate questions which are either too 'difficult' or too 'easy'. Such elimination can thus help strengthen the assessment process by achieving a mix of questions more appropriate to the level of challenge aspired to (pre-testing of questions has already been mentioned above). General problem areas within the module may also be identified and subsequent remedial action taken in regard to teaching and learning. In a similar manner, weak students can be identified within the module, as can students who are not being sufficiently challenged.

In addition to this use of statistical analysis to evaluate questions, feedback from students can be incorporated into every assessment, thus enabling the regular consideration of issues of concern raised by the student cohort.

TOP TIPS

Always pre-test questions and assessments in order to identify and eliminate weak questions.

2.9 Choosing your CAA system

This is a difficult issue given rapidly developing technology plus tightening cost control. Some points that should be borne in mind are given here.

The benefits of a customised system

There will always be benefits from adopting a system that is specifically designed for your own institution. This may allow the closest tailoring of your resources to your needs and syllabus. However, the downside of such an approach is often the need for very specialised IT staff capable of constructing this kind of authoring shell.

In addition, issues of maintenance and future development also need consideration, together with the questions of compatibility of any customised system with other standardised managed learning environments that the institution may be using now or may move to at a future time. In the Aberdeen case, there was no alternative to a customised system at the time that the project was first undertaken.

TOP TIPS

Before buying in a package, consult colleagues at other institutions (your LTSN co-ordinator can help).

Alternatives to customised systems

Some academic institutions may offer free or relatively cheap CAA systems that can be adapted to particular needs. Again, whether support is offered for the system and whether present support will continue into the future will be relevant, as will compatibility with other systems. In some cases (such as Clyde Virtual University: <http://cvu.strath.ac.uk/>), collaboration may exist between different establishments in providing online assessment engines that can be customised normally within certain limits. An important point that should always be borne in mind is that there are viable alternatives and real choices – not just the automatic choice of a commercial package (Crofts et al., 2000).

If customised systems are not possible and none of the non-commercial packages is suitable, then commercial packages can normally be found that will offer a varied range of possibilities within the economics discipline. WinEcon (now web based) claims to be the world's largest and most successful computer-based learning package for introductory economics and is certainly worth a look: <http://www.webecon.bris.ac.uk/winecon/>.

In addition to stand-alone packages such as WinEcon, an increasing number of managed learning environments, such as Blackboard (<http://www.blackboard.com/>) and WebCT (<http://www.webct.com/>), now offer the inbuilt facility for assessment as part of their overall package. However, unlike WinEcon, this normally needs to be created by those using the package – similar to Oracle-based systems.

2.10 Research and useful links

The following links may be found useful in allowing academics in the economics discipline to research further the particular issues that are relevant to their stage of adoption of CAA.

Perhaps the most useful individual site on all matters to do with CAA is: <http://caacentre.lboro.ac.uk/index.shtml>. Although sadly this is no longer maintained, it can still be accessed via a search on Google (using the cached pages) and is well worth a visit for a host of relevant materials. Hopefully some way may be found to keep this material available in future.

Recommended on the above site are the following printed articles, aimed at those beginning the road to using CAA:

'Using computers in assessment' (1997) in G. Brown with J. Bull and M. Pendlebury, *Assessing Student Learning in Higher Education*, Routledge, London, pp. 202–21.

D. Stephens and J. Bull (1998) 'Computer-assisted assessment: suggested guidelines for an institutional strategy', *Assessment and Evaluation in Higher Education*, vol. 23, no. 3, pp. 283–94.

J. Harvey and N. Mogeey (1999) 'Pragmatic issues when integrating technology into the assessment of students', in S. Brown, J. Bull and P. Race (eds), *Computer-assisted Assessment in Higher Education*, Kogan Page, London.
<http://caacentre.lboro.ac.uk/resources/faqs/fqgen9.shtml>

A very useful website containing links on all aspects of CAA, including case studies, is run by the Learning Technology Development Unit (LTDU):

http://www.herts.ac.uk/ltdu/learning/reading_list/caa_procon_read.html

Research into CAA specifically aimed at the economics discipline appears very limited, although an increasing amount of work is being done on generic issues related to CAA which are as relevant to economics as to other disciplines. A useful collection of papers – proceedings of the annual (and ongoing) Computer Assisted Assessment conferences held at Loughborough University – can be found at the following site:

<http://www.lboro.ac.uk/service/ltd/flicaa/conferences.html>

Some general points on computer-assisted assessment can be found on the LTSN site at: <http://www.ltsn.ac.uk/genericcentre>

Computer adaptive testing is dealt with in a very interesting article ‘This test is for you’, available from Wired magazine at: http://www.wired.com/wired/archive/3.01/adaptive_pr.html

The following references in this section have been drawn from work done at the Loughborough conferences.

Interesting work has been done by Crofts *et al.* (2000) examining the decision-making process followed in identifying how web-based assessment tools can complement a broader assessment strategy. Crofts states that ‘there are genuine choices to be made between [software] engines – rather than simply plumping for the popular commercial product’:

<http://www.lboro.ac.uk/service/ltd/flicaa/conf2000/pdfs/mcalpinem.pdf>

From a similar perspective, Sclater and Howie (2000) illustrate how Scottish universities are working together to develop a generic (i.e. not engine dependent) online assessment engine, which may be worth investigating:

<http://www.lboro.ac.uk/service/ltd/flicaa/conf2000/pdfs/sclatern.pdf>

Increasing attention is being paid to the matter of question banks, and how good ones may help pedagogy, while bad ones may hinder. Dalziel (2000) examines the trend of an increasing number of textbook publishers to bundle question banks with their product, if the textbook is adopted. This has dangers but also positive possibilities, and Dalziel examines current academic practice, which often involves the creation of a hybrid of questions from such banks, rewritten, customised and added to by lecturers:

<http://www.lboro.ac.uk/service/ltd/flicaa/conf2000/pdfs/dalzielj.pdf>

In an interesting paper, Davies (2001) outlines innovative work allowing computer-assisted peer assessment of essays (at levels 1, 2 and 3), which illustrates how CAA can move beyond lower-order skills towards developing higher-order skills:

<http://www.lboro.ac.uk/service/ltd/flicaa/conf2001/pdfs/e2.pdf>

Creating fruitful collaboration between different interest groups is necessary to ensure that a quality product is achieved which enhances education. White (2000) concentrates on the need for close collaboration between subject specialists, educational developers (i.e. academics who specialise in assessment methods) and those who have extensive experience of the use or implementation of CAA:

<http://www.lboro.ac.uk/service/ltd/flicaa/conf2000/pdfs/whites.pdf>

2.11 Examining available software

Below is a very brief selection of websites displaying some of the currently available software, which may illustrate feasible CAA systems:

WinEcon: <http://www.webecon.bris.ac.uk/winecon/>

Clyde Virtual University: <http://cvu.strath.ac.uk/>

The *Hot Potatoes* site (<http://web.uvic.ca/hrd/halfbaked>) includes six applications, enabling you to create interactive multiple-choice, short-answer, jumbled-sentence, crossword, matching/ordering and gap-fill exercises for the World Wide Web.

Examples of where objective testing has been used to assess higher- and lower-order skills are found on the following site (click on 'Try it Out'):
<http://www.questionmark.com>

Examples of more complex types of CAA questions (albeit from disciplines other than economics) may be found at:
<http://medweb.bham.ac.uk/caa>
<http://www.brookes.ac.uk/geology/8307/a40frame.html>

Blackboard (<http://www.blackboard.com/>) and WebCT (<http://www.webct.com/>) both offer increasing integration of CAA into their managed learning environment.

3 Case study: Macroeconomics at the University of Aberdeen (the 'Aberdeen case')

3.1 Background

The purpose of this case study is to describe the implementation and learning outcomes of computer-aided assessment and independent learning projects introduced into second-year macroeconomics teaching at the University of Aberdeen in 1999. Evaluation is benchmarked against the learning objectives (specified by the course, university and discipline) and compared with previous practice using traditional methods.

It is important when considering the introduction of any new and innovative practice to be aware of the nature of the target clientele and the institutional environment you operate in.

A project that works well in one location and in one set of circumstances may have to be intelligently modified to be able to work well in another. Thus, in section 3.2, the environment within which the project was devised is described together with how it informed the choice of learning and teaching methods developed. Section 3.3 discusses how the intended aims and learning outcomes were determined – what was to be achieved and how could it be done given the operational environment and the resource constraints faced? In section 3.4, the practical implementation of the project is described – what was done and when?

Finally, section 3.5 summarises the student experience of the project and evaluates the project from the point of view of the teaching team and the wider community. Whether the project fulfilled its ambitions or not, what improvements could be made, and prospects for the future are assessed.

3.2 The learning environment

Over 400 students take a course in economics in their first year at the University of Aberdeen, often as part of a programme leading to finance or management degrees as well as economics itself. Thus, given the low level of pre-exposure to economics and the diverse nature of the clientele, first-year courses are relatively general in nature and do not focus heavily on the technical training often found in English degree programmes in economics.

In the second year, approximately 100 students take second-year macroeconomics. Most of these will go on to take degrees in economic science, either as single honours or joint honours with other disciplines. Second-year modules must therefore make the bridge between a very general first year and the demands of honours-level study. The second year must provide a sufficiently rigorous training to enable those students who wish to cope successfully with the rigours of honours-level study to do so. For many students this implies a substantial step up from first to second year. This is particularly the case in connection with the transition from the mainly descriptive and diagrammatic analysis used in the first year to the more technical algebraic treatments required in the second year. One challenge, therefore, of developing teaching methods in this environment is to enable students to make this transition more easily.

3.3 Objectives and constraints

In this section what the objectives were and how this particular implementation of computer-aided assessment and independent learning helped to achieve them are explained. As many of the relationships in macroeconomics are complex and interrelated, a vital skill for any macroeconomist is to be proficient at model building and policy analysis. In the UK this centres on a modelling structure first devised in the early 1960s by Mundell and Fleming. This powerful model enables practitioners relatively easily to conduct policy analysis appropriate to low-inflation, open economies such as the UK, EU and USA.

However, to do so effectively requires repeated calculations of an algebraic nature. Thus, students face a double hurdle of not only attempting to master the intuition behind how government policy or external shocks affect the macro economy, but also developing the essential algebraic and other quantitative skills. The nature of these skills is such that they require repeated exposure and practise to master them, which cannot be effectively provided in a traditional lecture setting; nor would the provision of a sufficiently small group of tutorials be viable given resource constraints.

And so an innovative use of IT was called for that facilitated the use of independent learning strategies by students, repeatedly exposing them to model-building techniques, but taking some of the 'grind' out of the calculations so that they could focus more clearly on the underlying theory and policy analysis.

Furthermore, the QAA Economic Benchmark document states that the aim of degree programmes in economics should (among other things) be for students to: 'apply the knowledge and skills they have acquired to the solution of theoretical and applied problems', become equipped with 'appropriate tools of analysis to tackle issues and problems of economic policy' and have 'an ability to develop simplifying frameworks [and] appreciate what would be an appropriate level of abstraction'.

The model-building techniques and policy analysis developed in our project very clearly satisfy these key objectives. Moreover, the project clearly 'develops a range of transferable skills of value in employment' and the use of computing skills specified in the benchmark document. Thus, the project satisfies discipline benchmarks as well as course and university requirements.

Given the binding resource constraints faced, any potential solutions had to be very parsimonious in terms of both developmental and running costs. The first requirement was for powerful, but user-friendly computational software. The university had site licences for Microsoft Office 97 with plans to migrate to Office 2000. Thus there was assurance of software continuity in this respect, an important incentive when developing applications requiring it.

This particular suite of software includes Microsoft Excel, which provides substantial computational power and the ability to visualise results graphically in an easy-to-use Windows-based environment. It is also a package that is in use by the vast majority of small and large corporations around the world, and competence in its application was a transferable skill highly regarded by the university's Employers Liaison Group. Excel therefore stood out as the first choice in terms of maximum computational and graphics power at zero marginal cost.

The second requirement was for software to enable the delivery of web-based material to students (including questionnaires to evaluate teaching and learning) and to provide information to students (including course information and the results of assessments). The final requirement was a testing environment; software that would deliver the assessment, manage the online testing of students, and provide the results of the assessment to the assessor.

Packages such as WebCT and Blackboard could satisfy both of these requirements, in what is now termed a Virtual Learning Environment (VLE). However, in 1999 the university had only just begun to trial the use of WebCT, and its use required a steep learning curve in comparison to the more beginner-friendly Blackboard. On the other hand, the department had substantial experience in using traditional web-based media in conjunction with the university's Oracle-based student management information system (HEMIS). A number of departments had also begun to trial the use of Question Mark as their testing environment, with some degree of success. The university had acquired a site licence for the networked use of the testing environment, requiring departments only to purchase individual Designer and Reporter components (the Question Mark components that are used to design the assessments and manage the marking and reporting of results, respectively).

The decision was therefore taken, following extensive trials and consultation, to use Question Mark as the testing environment and traditional web-based media for information dissemination and student management.

The net cost of this was the Reporter and Designer modules of Question Mark. At the time, the university's Learning Technology Unit (LTU) was operating a small grants scheme to promote the innovative use of IT in teaching, learning and assessment.

Projects involved bidding for support time, the idea being that this support would enable innovative projects to be established, with the understanding that departments would maintain the projects in subsequent years. The department's bid was successful and was helped greatly by the LTU experts in CAA design, conducting useful evaluations (both web based and via focus groups), and writing scripts to automate the web integration.

3.4 Practical implementation

In this section we detail what was done in practical terms and the time scale involved. The second-year macroeconomics course takes place in the second semester of the academic year in Aberdeen. Learning facilitation takes place over a 12-week period commencing at the beginning of February. This is interrupted by the three-week spring holiday, dividing the 12-week teaching bloc into two blocs of eight and four weeks respectively.

In the first two weeks, students are entitled to 'shop around' and audit different courses, so it is important in this fluid environment that no material upon which assessment may be based is

delivered. It was also desired to retain, as part of the overall module assessment, an essay-based piece of work, as it was felt that this developed important academic skills.

The topic that most lent itself pedagogically to this mode of assessment was growth theory, which was taught in the final four weeks of the course. Thus, these timetabling constraints necessitated the independent learning and computer-aided assessment to be contained in weeks 3 to 8.

Weeks 3 and 4 were used (see the Gantt chart in Figure 1) to begin the delivery of the web-based material, which provided the background study material that we expected students to use in preparation for attempting the computer-based work. Students could download this material in PDF format (Microsoft Word was used to edit the documents and equations, and Jaws PDF creator to convert the document to PDF format). Great care was taken to ensure that the same notation was used in the lectures, tutorials, computer workshops and web-based materials as in the textbook (Gordon's *Macroeconomics*, 8th edn).

In weeks 5, 6 and 7, a series of three supervised computer workshops were provided. The first workshop hour was concerned exclusively with Excel skills. The task that students were given was to set up a basic worksheet that would enable them eventually to complete the assessment. They were given written documentation in the course guidebook on how to achieve this goal. It detailed a suggested structure of the worksheet and how to enter formulae (which was revision for them – see Box 6). The formulae were obtained from the web material and supporting lectures, but needed to be rearranged appropriately for the tasks at hand.

Thus, entering formulae was not purely a mechanical exercise and required students to think about what they were doing. The task was deliberately too large for the 'average' student to complete in the supervised hour, thus requiring private study. It was stated explicitly to students that they would need to take advantage of the 24-hour computer suites available.

Box 6 Information skills at Aberdeen

All students entering the Aberdeen MA programme are required to attend an information skills course in the first week of their first year. This provides students with benchmark skills in Microsoft Word, Eudora (e-mail) and Netscape (web), as well as library skills (via WebCT). They are pre- and post-evaluated on this course. Once the required skills are achieved, they are passed. Those with extensive IT skills already are often passed within the first day, though most manage to succeed within the first week. Extra classes are run in the following weeks for those who don't. The Senate Grid states that further skills, such as Excel and PowerPoint, may be embedded in courses, or acquired by dedicated courses run by DISS which now lead to the ECDL qualification. In economics, Excel skills are embedded within the year 1 microeconomics course (which has also been the subject of a CAA project: see www.abdn.ac.uk/diss/ltu/projects/econproj.hti). The second-year macroeconomics course therefore builds upon this provision.

In the second workshop hour, students were tasked with completing the trial assessment – a dummy run for the real thing, giving them experience of the testing environment, Question Mark. This trial assessment gave them feedback after each question so that they were able to assess how well they were doing, and how to correct any mistakes in their spreadsheet. This second supervised workshop also gave them the opportunity to consult their tutor on any questions arising from their private study in the preceding week.

The real assessment then had to be completed over the next week and a half in their own time. A catch-up lab was held the following week for any students who had been ill or otherwise absent, or just required extra help and advice.

A further supervised workshop was held using the package LiveEcon, an interactive suite of independent learning software under development by The Enterprise Library Limited (TELL). This was being simultaneously trialled at Aberdeen and the Judge Institute, Cambridge. Students were introduced to the modules relevant to the Mundell–Fleming model developed at Aberdeen, to complement their independent learning experiences thus far. Many students enjoyed having direct impact into the development of this package, since the program’s designer regularly attended the workshops to get feedback directly from the students.

Students completed the real assessment online in their own time. It was closely based on the trial assessment, but used different coefficient values and some question variation. For each question students were asked questions drawn randomly from a library of similar questions. This security feature meant that no two students could ever be asked the same sequence of questions. For additional security, the log-on time and terminal details of each student were recorded.

Once the deadline had been reached, the Reporter software was used to analyse the results. These were then processed and uploaded to the web, where students could access them together with a suggested solution, but only once they had completed an online evaluation questionnaire! This was mounted a few days before the spring break so that students could go on holiday knowing how they had done. In subsequent years, the turnaround time was reduced to as little as one day, meaning that the deadline could be pushed back and students given more time to complete the assessment.

The questions are written during November and December to ensure that, whatever may happen in January, the system is ready for use in February. In the first year of implementation, the independent learning material was also produced in this period. The required time decreased in subsequent years, as only minor corrections were needed to the web material, and the previous year’s test could be used as a ‘template’ for the current year.

	01		02															
	Nov.	Dec.	Jan.	Feb.				Mar.				Apr.			May			
Week no.				1	2	3	4	5	6	7	8	V	V	V	9	10	11	12
Question design																		
IL web material mounted																		
Workshops																		
Trial assessment mounted																		
Real assessment mounted																		
Results and evaluation																		
Traditional assessed essay																		

Figure 1 Gantt chart (2002)

3.5 Evaluating the experience

From a purely pedagogical viewpoint, the project has been a resounding success. Students’ understanding of the model used and the transmission mechanism of policy in the system has

vastly improved compared with previous practice. This is borne out in the final exam-based questions that cover these topics.

The mean CAA marks tend to be higher than for traditional essay-based coursework (which in turn are higher than for closed-book exams). There is also evidence of some degree of team working among some groups of students, though this is regarded as being a positive development.

After all, even if the input of each team member were unequal, a student doing most of the explaining would benefit from the enhanced knowledge of the subject gained from ‘teaching’ it, while those being ‘taught’ benefit from an alternative exposition and positive peer role models.

Focus group evaluation revealed that students found the set-up costs high in terms of work, effort and understanding, but the assessment relatively easy by comparison. They found that understanding how the model works and how to implement the system in Excel is harder than answering the questions about the system. This is also positive, since it demonstrates that students have been encouraged to do some harder ‘deep learning’, which in the long run they find more valuable. Comments like ‘the whole programme was a big boost in understanding the workings of macroeconomics’ and ‘everything was well organised’ sum up the general mood of the class.

Inevitably, a small number of technical problems were encountered, but all were resolved. Since the icons for the trial assessment (which students can do as many times as they like) and the real assessment (which they can only do once) are similar in appearance, some students were mistakenly beginning the real assessment and having to abort illegally. This was addressed in subsequent years by restricting the mount time and inserting alert screens. A desire for more Excel guidance was catered for by increasing vastly the content of the course booklet. A summary of the problems and evaluations are found at: www.abdn.ac.uk/diss/ltu/projects/econ2proj.hti

3.6 Final advice

Finally, Box 7 contains a few words of advice to those considering implementing a similar project.

Box 7 Tips on implementing a CAA system

- First, design the aims and objectives of the project for specific pedagogic reasons, appropriate to the conditions in your department and university. Ask how it will enhance students’ learning, rather than how much time or resources it will save. At the same time, CAA, though it may cut the marking load overall, rarely saves, and often adds to, the developer or co-ordinator’s time.
- Second, make sure that your aims are achievable given the resource constraints and support available – it is better to do it well on a small scale at first, than badly on a large scale all at once.
- Third, conduct meaningful evaluations measuring how well the objectives have been achieved in the appropriate timescale. Verbal feedback through focus groups, for example, can be tremendously useful – tutors should be flexible enough to respond to feedback both during and after the implementation.
- Finally, do not forget to explain to your students why CAA is being implemented and at the same time involve them in it. This way they will share in the project’s objectives.

4 Where next?

It is clear that beginning from its original use in automatically assessing objective questions, the principles of computer-aided assessment are now speedily being applied to ever wider areas of assessment and to challenges such as essay marking and higher-order skills. Like other tools in one's pedagogical armoury, the issue of CAA cannot be ignored even in the most traditional of establishments. Practitioners in economics therefore have little option but to engage with the existence of CAA and examine how and where it can become a true benefit, rather than a cost-saving 'quick fix'.

In addition, and to an increasing degree, greater use and availability of IT and the internet is undoubtedly affecting the general response of higher educational establishments to how education is delivered. It is very likely that the future student may have a greater choice of where she or he studies and the manner in which that study takes place. It is likely that online learning and teaching will play an ever-increasing role in the delivery of education. The development of CAA moves with the grain of these changes and is in tune with such development. The increasing demands from students for this online option will inevitably lead to a greater demand on academics themselves to add proficiency in this area of pedagogy to the other areas that they have developed in the past. The challenge will be to maintain the integrity of what can be offered and achieve the right balance between distance and automated learning and face-to-face encounter.

References

Brown, G., with Bull, J. and Pendlebury, M. (1997) 'Using computers in assessment', in *Assessing Student Learning in Higher Education*, Routledge, London.

Crofts, A. (1999) *Enabling Re-use of CAA by Design*, Third Annual Computer Assisted Assessment Conference, Loughborough.

Crofts, A., Foster M., et al. (2000) *User Experiences of Choosing and Implementing Web-based Computer Aided Assessment*, Fourth Annual Computer Assisted Assessment Conference, Loughborough.

Dalziel, J. (2000) *Integrating Computer Assisted Assessment with Textbooks and Question Banks: Options for Enhancing Learning*, Fourth Annual Computer Assisted Assessment Conference, Loughborough.

Davies, P. (2001) *Computer Aided Assessment must be more than Multiple Choice Tests for it to be Academically Credible*, Fifth Annual Computer Assisted Assessment Conference, Loughborough.

Dowsing, R. D. and Long, S. (1999) *Evidence, Assessment Criteria and the Difficulty of Automated IT Skills Assessment*, Third Annual Computer Assisted Assessment Conference, Loughborough.

Haladyna, T. M. and Downing, A. (1989) 'A taxonomy of multiple choice item writing rules', *Applied Measurement in Education*, pp. 61–86.

Haladyna, T. M. (1994) *Developing and Validating Multiple Choice Questions*, Lawrence Erlbaum Associates Inc., Mahwah, NJ.

Harvey, J. and Moge, N. (1999) 'Pragmatic issues when integrating technology into the assessment of students', in S. Brown, J. Bull and P. Race (eds), *Computer-assisted Assessment in Higher Education*, Kogan Page, London.

Heard, S., Nicol, J., *et al.* (1997) 'Setting effective objective tests – a guide to creating effective objective paper- and computer-based tests', University of Aberdeen.

Hill, S. (2000) 'Guidelines for using technology in examinations', <http://www.disinhe.ac.uk/library/article.asp?id=33>.

McKendree, J. (2002) 'Negative marking and gender bias', CAAC, http://caacentre.lboro.ac.uk/resources/faqs/negative_marking1.shtml.

McKenna, C. and Bull, J. (1999) *Designing Effective Objective Test Questions: An Introductory Workshop*, Third Annual Computer Assisted Assessment Conference, Loughborough.

Mulligan, B. (1999) *Pilot Study on the Impact of Frequently Computerised Assessment on Student Work Rates*, Third Annual Computer Assisted Assessment Conference, Loughborough.

Nicholls, J. (1999). *The Web between Lectures and Self-managed Learning*, Third Annual Computer Assisted Assessment Conference, Loughborough.

Slater, N. and Howie, K. (2000) *The Scottish Computer Assisted Assessment Network*, Fourth Annual Computer Assisted Assessment Conference, Loughborough.

Stephens, D. and Bull, J. (1998) 'Computer-assisted assessment: suggested guidelines for an institutional strategy', *Assessment and Evaluation in Higher Education*, vol. 23, no. 3, pp. 283–94.

Whigham, D. and Houston, J. (1999) *On-line Delivery and Marking of Excel-based Assignments*, Third Annual Computer Assisted Assessment Conference, Loughborough.

White, S. (2000). *Creating Large-scale Test Banks: A Briefing for Participative Discussion of Issues and Agendas*, Fourth Annual Computer Assisted Assessment Conference, Loughborough.

Zakrzewski, S. (2002) *Implementation of Large-scale, Computer-based Exams*, Computer Assisted Assessment Centre, Luton. <http://www.icbl.hw.ac.uk/ltdi/assessit/luton.htm>