COMPUTER AIDED ASSESSMENT MUST BE MORE THAN MULTIPLE-CHOICE TESTS FOR IT TO BE ACADEMICALLY CREDIBLE?

Phil Davies

Computer Aided Assessment MUST be more than multiple-choice tests for it to be academically credible?

Phil Davies
School of Computing
University of Glamorgan
Treforest
Pontypridd
Mid Glamorgan
South Wales
CF37 1DL
pdavies@glam.ac.uk

Abstract

Computer Aided Assessment (CAA) is being proposed as the means of providing formative testing to the ever-increasing numbers of students involved in higher education. A large proportion of this testing is being based around the use of objective based multiple-choice tests. These tend to make use of the process of select one out of three possible answers, or to improve the reliability of the process, one out of four! Obviously there are alternatives to the above, however, a lecturer initially will use the method that is easiest for them. With the complexity, multi-functionality and hence steep learning curve involved in setting up these systems, many lecturers are dumping them before their benefits can be achieved. Question banks are available, however, these normally require an initial financial outlay. The questions offered are often general, and may not directly map to the areas covered in a particular module. No matter what system is utilised for these objective tests, the criticism offered by CAA sceptics is "we are not developing employable and transferable skills". Students on leaving higher education will rarely be expected to produce answers for their employers that require multiple-choice skills, but will be expected to produce reports, presentations, etc. The question that springs to mind is ... what is our job as educators, to produce clones, or to develop and nurture broader skills? The use of formative multiple choice tests is also causing a problem in that students are not being "prepared" to sit their final examinations. These often still take the format of select three from five questions and then write essays for each of the selected questions in a limited time period. Having students develop essays as part of their formative/summative assessment throughout the course of a module, again brings us back to the time consuming problem of having to mark and provide formative feedback.

This paper introduces the audience to the Computerised Assessment and Plagiarism system (Davies 2000), that provides an on-line means of students assessing the essays of their peers, and providing formative feedback. This system has been successfully used at levels one, two and three of an undergraduate programme in

the field of computer studies at the University of Glamorgan. It has been used for continual assessment at level one, a combination of multiple choice / peer assessment at level two, and for self, peer and reflective self-assessment at level three. The use of this networked tool has produced major positive benefits both for the students and staff. Its acceptance has not only provided an efficient method for formative / summative assessment, but has also aided in developing the students' essay writing skills. From a lecturer's point of view, those who in the past have been sceptical of the use of peer assessment and the more general use of objective testing, have become much more receptive to the introduction of these innovative assessment methods making use of CAA.

Introduction

Multiple Choice Questioning via the use of computer systems can no longer be thought of as innovative and is being used more and more in education " … over 70 universities and colleges of education … CAA is overwhelmingly used for summative purposes" (CAA 1999). Many tried and tested systems are in the market place, both commercial and those developed by various academic institutions. These systems have been widely used and major benefits both educational and from reducing the lecturer-marking load have been achieved (Davies 1999; Thelwall 1998; Sly 1999). However, it should never be forgotten that " .. the most important thing we do for our students is to assess them" (Race 1995), and we must always be looking at why we are using a particular form of assessment with respect to the benefit accrued by the students. The reasoning of using computerised multiple choice tests because it makes it easier for lecturers to accomplish the often time consuming and tedious process of marking is not acceptable. Multiple choice questioning has not met with total acceptance, both for reasons of academic acceptance and also with respect to their unfairness towards gender and certain sections of society (Childs 1990). This paper is not condemning the use of multiple choice questioning, but is suggesting that we need to be addressing additional needs of our students through the assessment process. It should be noted that "..the development and integration of Computer Aided Assessment has been done in an ad hoc manner" (McKenna & Bull 2000). Commercial systems are not easily integrated into existing domains without specialist support and assistance being required, often resulting in a substantial initial cost.

Rather than seeking to replace traditional assessment with multiple choice, or not using multiple choice at all, we should be attempting to integrate different forms of computerised assessment in a manner allowing us to assess a wider range of student skills, rather than just objective testing. This paper will briefly describe a system of computerised peer assessment of essays that has been used successfully at levels one, two, and three of an undergraduate programme (Davies 2000).

Assessment Needs

Before looking at the methods of assessment available, it is important to be clear about the skills that we are trying to assess. There still appears to be scepticism in the acceptance of multiple-choice questioning a means of being able to assess the

higher level skills of analysis, synthesis and evaluation (Bloom 1956). Numerous papers have been produced attempting to support the use of multiple-choice testing for attaining higher level skills and also attempting to correlate multiple choice with examination performance. These have been equally met by a barrage of papers presenting the counter argument “a 60 question, four choice test is inherently too unreliable for the demands that are commonly placed on it” (Burton 2001).

What we are seeking is a method of assessment that we can trust to provide quality in the assessment process. The extreme example below illustrates a scenario where multiple-choice testing falls short of these requirements:

Example: Three students are to be asked one question to decide if they are to be offered a job. The question being “what colour is the sky?”

Student A enters the interview and is asked the question. The student asks could you provide me with three alternative answers … the questioner provides one correct and two distracters blue / green /pink… the student is still unsure but guesses blue.
Student B is asked the same question and gives the answer blue.
Student C is asked the same question, and answers that the sky will be different colours depending upon the weather, time of day, etc., but normally will be blue.

If the employer only had one pass or fail box available, then each of the candidates would have passed the assessment test.

Student C in the above example showed skills of analysis in coming to the final answer, yet this would not have shown in the results. An essay would have provided a far more judgmental means of selecting the best future employee. “Essays allow for student individuality and expression” (Race 1995). However, it should also be noted that “essays are demonstrably the form of assessment where the dangers of subjective marking are greatest” (Race 1995). What is required is a means of assessing these essays in an objective manner. Free text response systems tend to rely on “a bounded spectrum of content” (Christie 1999), which works against our needs of assessing higher levels of understanding. Currently there is no readily available free text marking system “…. regarded as the Holy Grail of computer aided assessment” (Whittington & Hunt 1999).

The system outlined below attempts to provide a solution to the previous problems. It makes use of computerised peer assessment to provide the marking of student essays, hence providing assessment of essays, but not increasing the marking load on the lecturer. Full details concerning the system and its use can be found at (Davies 2000)

Computerised Assessment by Peers (with Plagiarism)

Peer assessment is not a new method for assessing students. (Dochy, Segers & Sluijsman 1999; Mowl & Pain 1995; Boud, Cohen & Sampson 1999). However, the use computerised peer assessment is still in its infancy (Robinson 1999). Most of the studies concerning peer assessment suggest that there are positive benefits to the

student in undertaking the process of marking their peer's work " .. pushes the students to think about the qualities that contribute to good work" (Robinson 1999).

The feedback from the use of the CAP system has been extremely positive in the areas of:
Learning benefits
Copying of good practice
Identification of poor practice
Immediate direct feedback
Reduction in plagiarism
Student and Staff acceptance
The negative aspects of the feedback related to:
Conflicting peer feedback occasionally being presented to a student
Occasional "wild" mark

The problem with any peer assessment operation is how to get the students to perform the marking process in a professional and objective manner. This has been overcome in this process by the awarding of marks for the marking process. These have been based upon the students' being able to justify their marks in the comments provided. This has resulted in the lecturer's marking being restricted to marking the "marking process" rather than the essays themselves.

As previously stated in this paper, multiple-choice questioning can be integrated into the assessment process if it benefits the students. Included in the assessment process either side of performing the peer marking process, the students undertook a multiple-choice test. The statistical results showing the benefits in undergoing this process are fully reported in (Davies 2000).

The student feedback reported two interesting comments
"over 75% of the students made the point that they had put in a considerable amount of extra effort into developing the report, due to the fact that they were to be further tested on their knowledge" (making use of the multiple choice tests),
and also
"the writing of a report is just the rephrasing of web source, the tests checked my understanding" (Davies 2000).

The use of an integrated approach to the assessment process has provided a method of testing both objective and subjective knowledge within the subject area.

The CAP system has been further used as part of the assessment process at level three of the undergraduate programme. It has been used as a method of providing self-, peer- and reflective self- assessment. This has produced an excellent integrated assessment process that has produced results that indicate that nearly 80% the students assessed were able to produce a self assessment mark within 10% of the resultant peer mark produced.

Finally the system has been used within level one of the programme, where the students have been continually peer assessed throughout a module for formative

purposes, with the final essay being an amalgamation of the previously peer assessed assessments.

Conclusions

The use of multiple-choice questioning is often of a formative nature. Often the systems merely provide the correct answer, without any additional explanations. With the increased number of students involved in higher education, computerised multiple choice can provide an excellent support tool, possibly even replacing the need for tutorial support. What we need are "***explainative***" systems that can be a mix of assessment and Computer Aided Learning.

Computerised peer assessment via the CAP system has provided major benefits both for the students and the staff. It has provided a means of computerised assessment that allows the production of essays, and thus offers a more acceptable method for assessing the higher level skills required within higher education. The integration of multiple choice and essays within the assessment process has produced a reliable and accepted system both from the student and lecturer perspective.

The use of the system for continuous assessment at level one has resulted in divergent rather than convergent education. This may indicate that the methods used for computerised assessment need to employ methods of adaptive testing to ensure that the brighter students are academically stretched via the assessment process.

Overall, if we are able to assess both subjective and objective ability in a module's assessment process via computerised means, then issues such as acceptance and quality can be overcome.

References

Bloom, B.S. (1956), *Taxonomy of Educational Objectives: Handbook of Cognitive Domain*, New York McKay.

Boud, D., Cohen, R., Sampson, J. (1999), *Peer Learning and Assessment*, Assessment and Evaluation in Higher Education vol 24 no. 4 pp 413-426.

Burton, R (2001), *Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answe*rs, Assessment and Evaluation in Higher Education, vol 26, no. 1,.

CAA(1999), The Computer Aided Assessment Centre http://caacentre.ac.uk/ .

Childs, R.A. (1990), *Gender Bias and Fairness*, Practical Assessment, Research and Evaluation, vol 2 no.3.

Christie, James (1999), *Automated essay marking – for style and content,* Proceedings of the 3rd Annual CAA Conference, Loughborough, pp 39-48.

Davies, Phil (1999), *Learning through assessment OLAL … On-line Assessment and Learning*, Proceedings of the 3rd Annual CAA Conference, Loughborough, pp 75-88.

Davies, Phil (2000), *Computerized Peer Assessment,* Innovations in Education and Training International, vol 37 no. 4 pp 346-355.

Dochy, F., Segers, M. , Sluijsmans, D. (1999), *The use of self-, peer-, co-assessment in higher education: a review*, Studies in Higher Education vol 24, number 3, pp331-350.

McKenna, C. and Bull, J. (2000), *Quality assurance of computer-assisted assessment: practical and strategic issues*, Quality Assurance in Education, Vol 8, number 1, pp24-31.

Mowl, G. and Pain, R. (1995), *Using self and peer assessment to improve students' writing: a case study from Geography*, Innovations in Education and Training International pp 324-335.

RACE. P. (1995), *Art of Assessing*, The New Academic 4,3 & 5,1.

Robinson, J.M. (1999), *Computer Assisted Peer-Review*, Computer-Assisted Assessment in Higher Education SEDA Kogan Press pp95-102.

Sly, L. (1999), *Practice tests as formative assessment improve student performance on computer-managed learning assessments* , Assessment and Evaluation in Higher Education vol 24 no.3 pp 339- 343.

Thelwall, M A (1998), *A unique style of computer-assisted assessment*, ALT J pp49-57.

Topping, K (1998*), Peer assessment between students in colleges and universities,* Review of Educational Research vol 68 no.3 pp 249-276.

Whittington, D. and Hunt, H. (1999*), Approaches to the computerized assessment of free text responses*, Proceedings of the 3rd Annual CAA Conference, Loughborough, pp 207-219.