Creating large-scale test banks: a briefing for participative discussion of issues and agendas

Su White, Hugh Davis,
University of Southampton
Department of Electronics and Computer Science
University of Southampton
Southampton, S017 1BJ

saw@ecs.soton.ac.uk
tel: +44 (0)2380 594471
fax: +44 (0)2380 592865

Abstract

This participative discussion will focus on the methodologies to be used for the production of a large-scale data bank of questions for use in all forms of assessment in higher education. This paper outlines some of the social and technical issues that the discussion will consider. It covers both the problems of building such a resource, and the rationale and methods of providing test bank data for a wide range of question delivery tools and methods.

Introduction

The University of Southampton has recently established a Learning and Teaching Task Force (http://www.ecs.soton.ac.uk/engfaculty/) in the Faculty of Engineering and Applied Science.

One early objective of the Task Force is to explore and increase the use of automated assessment for self-assessment, formative assessment and summative assessment - initially in the Department of Electronics and Computer Science. A key part of this activity will be to build up banks of peer reviewed test questions.

At the same time the Task Force team have recently (May 2000) been successful in receiving funding to establish a national Electrical and Electronic Engineering Assessment Network (E3AN - http://www.ecs.soton.ac.uk/E3AN/) supported via phase 3 of the Fund for the Development of Teaching and Learning. The project partners are a south coast consortium of Bournemouth University, Portsmouth University, Southampton Institute and the University of Southampton. The project will identify and disseminate good practice in all areas of assessment, but a major driver to this activity will be the creation of a nation-wide network of academic assessment consultants who will work in teams to create and peer review test banks of questions.

In order to provide a centre for this discussion we will focus on the issues that the E3AN project team is currently facing. Many of the agenda items to be addressed by the network are issues which have been presented and explored at previous CAA conferences and have been the subject for ongoing debate on various UK academic mailing lists.

This open discussion/networking session will explore:

- the use and application of test banks in automated assessment;
- the creation of peer reviewed test banks in engineering;
- the associated issues of test bank interoperability.

These are not new issues, but are worthy of ongoing consideration and discussion. Although our particular interest focuses on the engineering curriculum, many of the issues under consideration are of general interest. We will divide the discussion into two sections, the first dealing with the socio-technical questions, and the second dealing with the interoperability problem.

The purpose of this paper is to sketch out the lines of the existing debate prior to the focus group meeting in order to stimulate a more meaningful debate during the conference. A summary of the discussion will be posted on the project web site after the conference http://www.ecs.soton.ac.uk/E3AN/ .

The Socio-technical debate

It takes a long time to produce good questions for CAA, and the higher education community is very much aware of the problem of getting one teacher to use resources produced by another. In this section we will discuss the problems of how to create a test bank:

I. How to choose useful areas;

II. Which areas will offer the greatest gains;

III. Which areas lend themselves to CAA, and which don't;

IV. Creating test banks in a cost-effective manner;

V. How to classify questions for skills level, level of difficulty, cognitive skills level, and time required

VI. Subdividing question banks into pools

VII. How to establish an effective peer review process;

VIII. Barriers and drivers for use and dissemination;

IX. Distribution models, continuation, growth and maintenance of the test bank.

The interoperability context

There is a wide and diverse range of assessment engines in use in UK Higher Education and in the commercial sector, both in forms of specialised CAA tools and as assessment engines within virtual learning environments. No amount of investment in creating test banks can be of any great use unless those banks are available in a format or range of formats which can be used by the widest possible variety of different delivery platforms which have been selected by various target users. Discussion of interoperability issues has been brought into sharper focus by drives for web based assessment engines, and specialised mark up language followed by developments of XML and the related activities of the Instructional Management System (IMS) Test

Interoperability Standards has increased the possibility of exchanging questions between systems. In this section we will discuss such issues as:

I. Interoperability standards - reality of vapourware?

II. How should authors produce questions for maximum interoperability?

III. Are some question types more Interoperable than others?

IV. How do we choose a system for interoperability?

Some Socio-technical issues

Assessment in all its forms is a key part of Higher Education processes. Assessment shapes and drives students' learning[1]. Assessment must test the integration of appropriate knowledge and skills in a meaningful way. In order to facilitate student acquisition of these appropriate skills and abilities, effective teaching and assessment processes are necessary to promote optimal learning[2]. For these reasons there is a strong imperative to innovate assessment methods.

From the academic's point of view, the design, delivery, marking and moderation of summative assessments and coursework take up large chunks of a teaching workload, which must be repeated on a regular basis. The additional work needed to produce and mark diagnostic and formative assessments means that in some cases such assessments are reduced, avoided, or abandoned in the face of increasing student numbers.

These factors have been influential in the trend towards automated assessment. One significant area of challenges when implementing technological solutions to educational innovations is in achieving the correct balance between pedagogy and technology[345]. The early adopters of technological solutions are likely to be enthusiasts motivated by personal affinities with technology[6]. However it may be that some academics who have a strong insight into what constitutes effective and appropriate assessments fall into the category of individuals who are reluctant or even hostile to any move to computer based methods.

It has been suggested that the solution to such problems can lie in establishing multi skilled teams who work collaboratively to implement and embed learning technologies. The issue there is how in reality can we establish multi skilled teams? If we are to see any large scale transition to the introduction of new methods, where will we find sufficient numbers of individuals to make up these multi skilled teams?

In addition, a recurring theme in discussion concentrating on the enhancement of university level education returns to the debate of how we can develop and assess higher level cognitive skills[78]. Although there is on-going work in this area, the question remains to what extent questions and test banks in common use actually effectively deal with higher level cognitive skills.

There is a similar need for input from multi skilled teams when it comes to deciding how to choose useful areas for creating test banks. If we accept that pedagogy needs to be the leading force in test bank design, then we have to

accept that the decisions on what areas of the syllabus need to be addressed are best answered by academic subject specialists. But alongside those considerations, there is knowledge and experience of assessment methods which may be best understood by those with an educational developer perspective. Such individuals may or may not also be subject experts. In addition those with extensive experience of the use or implementation of CAA, whether from an academic or technical perspective, also need to be taken into consideration.

By taking into account the views from these different perspectives we will be able to identify which areas will perhaps offer the greatest gains. But when it comes to the question of how do we measure such gains the picture becomes more complex. What are the criteria on which we are going to measure potential gains? If we are concerned with saving time, then is that time which is already being used at very low cost, such as the long hours into the night which many academics spend on the regular tyranny of marking associated with end of semester examinations? Or is it saving time to do assessments such as ongoing formative assessments that have in many cases already been cut from the academic workload? Will the best savings come from dealing with very difficult assessment issues, which are perhaps currently avoided because they are problematic? Is using question banks to populate suites of self-assessment tests the most effective method? Can the questions actually be authored to some extent by the learners, and then captured for future re-use thereby making multiple savings?

The question of which areas lend themselves to CAA, and which do not is one which may have either a technical or subject specific response. From a technical perspective, it is perhaps related to cost saving issues as discussed above. From a subject specific view, in the context of engineering, matters that need to be tackled include identifying what sort of teaching takes place in engineering.

It is probably the academic subject specialists who have the best perspective on how to classify questions for skills level, level of difficulty and cognitive skills level, which they are testing and time required for the tests. To some extent some of these considerations will have been determined by external influences such as subject benchmark standards and requirements from professional bodies. However once again, there may be useful contributions on the practicalities of methods of classification from CAA specialists and educational developers.

The question of subdividing question banks into pools is to some extent an elaboration of the debate surrounding the selection of topic areas and levels for testing. However there are some additional issues which need to be explored if sets of like questions are to be placed in a pool for use by some random test generator. How can we be sure that generated tests will be fair to students?

Another area which draws on prior experience is identifying ways of creating test banks in a cost-effective manner. There has been extensive experience of creating test banks in some subject areas such as medicine. Are there quick and easy ways which can be shown to work? Every system seems to come with its own test creation wizard, on which its advocates heap great

praise, while novice users relate horror stories of a day's work creating a single question. Here surely lies a very strong driver to test bank interoperability.

The identification and use of appropriate time saving methods are also objectives in creating an effective peer review process. The E3AN team agreed to work together piloting the creation and review of a test bank so that they would experience at first hand the processes which they were proposing for their academic consultants. The project sees the peer review of questions as being essential in terms of achieving ownership of any test banks across the community. It also believes that the process is essential in terms of assisting the establishment of clear academic standards associated with the test banks. At the same time the approaches adopted need to make the best use of time, otherwise academics will be unable to participate.

Creating and reviewing the test banks is the first step in achieving their use and the dissemination of their use. Those who deal with embedding technology for learning and teaching are happy to acknowledge that a large part of the approaches that they adopt are concerned with managing change. This will be true whether they are operating as a project, e.g. a nationally funded activity such as the national CAA centre based at Luton, or whether they are institutional learning technologists or educational developers.

If the project is successful in dissemination and use of the test banks, then issues associated with identifying an appropriate distribution model will become important. The project is faced with the question of ensuring effective continuation, growth and maintenance of the test bank. Possible distribution modes include a cost free copyleft agreement, a system of microcharges, or via some kind of membership organisation run in collaboration with the LTSN Engineering or a professional body such as the IEE. We would be interested to hear views and experiences of successful models.

Some interoperability issues

There is a wide range of commonly used platforms in UK Higher Education which are specifically designed for computer assisted assessment, and there are frequent examples at this conference of new customised tools being developed by academics. Major activity in the area of CAA funded by the Joint Information Systems Committee (JISC) Technology Application Programme (JTAP) has supported development of customised tools at Leicester and Bristol. Bristol and Loughborough are both Universities with large and well known engineering departments who have a major interest in CAA. The Scottish Funding Council has invested in the Clyde Virtual University (CVU) which developed an assessment tool which is in use in a number of Scottish Universities.

In the commercial engineering world companies such as Microsoft, SUN and CISCO have extensive training programmes which administer large amounts of automated assessment with customised programmes using large-scale test banks.

In addition in both Higher and Further Education, and amongst private training providers there is now a move towards the adoption of a wide range of virtual

learning environments each with their own customised assessment engines as well as sometimes the capacity to interface with specialist engines. The investment of time and effort in establishing use of given platforms whether on a course, departmental, or institution wide basis often results in a high level of personal or institutional commitment to the chosen platform.

However no amount of investment in creating test banks can be of any widespread use unless those banks are available in a format or range of formats which can be used by the widest possible variety of different delivery platforms which have been selected by various HE institutions.

It would seem that the recent increase in activity on interoperability standards driven by the IMS[9], and the associated interest and debate on these issues would flag the dawning of a brave new world in test bank interoperability. But one question which needs to be asked is how relevant are the emerging standards to many of the systems actually in use across the sector? Is the reality that storing data in a comma separated variable file is actually the most realistic, lowest common denominator solution?

In addition, there are many institutions that do not make use of automated assessment in any widespread or systematic way. Despite advances in the use of automated assessment it remains true that the majority of assessment experienced by learners in UK Higher Education is paper based. Where test banks already exist, it is most likely that very many of these are paper based. For these reasons there is a strong argument for developing testbanks which can also be used in a conventional paper based context as well as by CAA programs and virtual learning environments.

Again perhaps simple formats may be the most effective, and a great deal of time can be saved in assembling test banks by ignoring the more esoteric, although educationally important specialised question types because realistically they are not actually used by very many people.

References

[1] Gibbs, G. (1995) Improving student learning through assessment and evaluation, Oxford Centre for Staff Development, Oxford Brookes University, Oxford

[2] Ramsden, P. (1992) Learning to Teach in Higher Education, Routledge, London

[3] Hall, W. and White, S. (1997) In Facing Up to Radical Change in Universities and Colleges (Eds, Armstrong, S., Thompson G and S, Brown.) Kogan Page, London, pp. 18-28

[4] Hall, W., Woolf, B. P. and White, S. (1999) Interactive Systems For Learning And Teaching, In Handbook of Internet and Multimedia Systems and Applications(Ed, Furht, B.) CRC Press, Boca Raton, Fla., pp. 375-398

[5] Hutchings, G., Davis, H., Hall, W. and White, S. (1994) Resource-based-learning: creating re-usable hypermedia for education In Media-Active Harnessing Multimedia in Higher Education, Liverpool John Moores University.

[6] White, S. and Maier, P. (1998) In Bringing Information Technology to Education (BITE), Vol. 1 University of Maastricht, Maastricht, pp. 117-123

[7] White, S. (1994) Apprentices of a new wave sorcery, Times Higher Educational Supplement.July 1994

[8] Bourner, T. and Flowers, S. (1997) Reflections on Higher Education, 9

[9] http://www.ims.org