Using evaluation to inform the development of a user-focused assessment engine

Conference or Workshop Item

oro.open.ac.uk

USING EVALUATION TO INFORM THE DEVELOPMENT OF A USER-FOCUSED ASSESSMENT ENGINE

Gráinne Conole and Niall Sclater

Using Evaluation to Inform the Development of a User-Focused Assessment Engine

Gráinne Conole
University of Southampton, Southampton, UK,
g.c.conole@soton.ac.uk

Niall Sclater
University of Strathclyde, Glasgow, Scotland
niall.sclater@strath.ac.uk

Abstract

This paper reports on the evaluation of a new assessment system, Technologies for Online Interoperability (TOIA). TOIA was built from a user-focussed specification of an assessment system. The formative evaluation of the project complemented this initial specification by ensuring that user feedback on the development and use of the system was iteratively fed back into the development process.

The paper begins by summarising some of the key barriers and enablers to the use of assessment systems and the uptake of Computer-Assisted Assessment (CAA). It goes on to provide a critique of the impact of technology on assessment and considers whether innovative uses of information and communication technology (ICT) might result in new e-pedagogies and practices in assessment. The paper then reports on the findings of the TOIA evaluation and discusses how these were used to inform the development of the system.

Introduction

Evaluation studies of the use of computer-assisted assessment (CAA) consistently reveal a set of enablers and barriers to the uptake and use of CAA in Higher Education (Table One). (Conole and Warburton, 2005) There is the attraction in terms of using CAA in terms of potential time saving and in particular time saved on teacher marking but this is coupled with the upfront investment of effort of writing the questions. Furthermore good question writing is a skilled technique and takes time to develop.

The use of shared question banks across subject domains has become increasingly popular particularly in science and engineering (Sclater, 2004). Interestingly one of the perceived benefits of CAA from the student perspective is the idea that somehow a computer marked test is more

objective. A potential drawback for some is the view that CAA is too restrictive and cannot be used effectively to measure higher order thinking. A national survey of CAA revealed that on the whole it was being used to support formative assessment (Warburton and Conole, 2003); the lack of widespread use of CAA for summative assessment is almost certainly associated with the perceived risks and security issues. Finally CAA users complain that many of the CAA tools themselves are still very rudimentary and restrictive.

Table One – Barriers and enablers to the uptake and use of CAA	
Enablers	Barriers
Potential time savings, especially with marking New pedagogical models Repurposing year on year Reflection on practice Shared question banks 'More objective'	Considered restricting Time and effort in development and management Difficult to measure higher order thinking Security issues Stress! Tools still rudimentary

ICTs offer a range of new 'affordances' (Conole and Dyke, 2004) which raise fundamental questions about how students should be assessed. Firstly, the Internet now offers students access to vast amounts of information and readily accessible resources, which raises questions about whether it is still therefore appropriate to focus assessment processes primarily on knowledge recall. Furthermore this has led to new opportunities for students to plagiarise through the use of online essay and assessment banks (Carroll and Appleton, 2001). This suggests that we should be shifting from an information-focused approach to learning to problem-based learning and inquiry. This requires a shift from assessing the 'products' of learning to concentrating on the processes of learning (de Freitas and Mayes, 2004; Beetham, 2004).

Similarly the new forms of communication and collaborations which are now possible with technologies also raise a number of issues about what we should be assessing and how to assess it – how do we measure the interactions which occur within an online discussion forum and how can we assess individual student contributions? What about those who do not contribute – the "lurkers" – are they opting out or learning differently – ie vicariously by reading and reflection on the postings of others? (Jones, 1999).

Specifying the roles and functions of an online assessment system

A number of different approaches can be taken to modelling complete "monolithic" online assessment systems. The various components of the system such as authoring, delivery and reporting can be detailed. Use cases (eg for a candidate taking a test or a marker marking an essay) can help to show how real users utilise different parts of the system. Sclater and Howie (2003) take a role-based approach and describe twenty-one user roles which can be combined in various ways as appropriate for "real" users. Thus a

departmental secretary might be a "scheduler", and a "results viewer", while a lecturer might be an "item creator", a "test creator", a "test viewer" and a "results viewer". In the "ultimate" assessment system an institution would be able to call these "real" people by names appropriate to their institutional context and to combine the low level roles in different ways.

The TOIA project

Understanding of the barriers and the enablers to the uptake and use of CAA, coupled with Sclater and Howie's articulation of the roles and functions of an online assessment system were used as a foundation for the development of the TOIA system. Funded by the Joint Information Systems Committee (JISC), TOIA is a freely available assessment tool which aims to guide practitioners through the process of developing and managing online assessments (www.toia.ac.uk). In addition the team undertook detailed formative evaluation studies during the development phase of the engine iteratively to improve and tailor requirements to end user needs. This paper provides a report on the evaluation systems and a critique of how this informed the subsequent development of the system.

Methodology

The aims of the usability and evaluation trials were to:

- test out the functionality of the TOIA system
- assess navigational and usability issues
- gain an understanding of the ways in which TOIA would be used
- identify any barriers to or enablers for CAA and in particular the use of TOIA.

The first phase involved usability trials of the TOIA prototype coupled with a broader evaluation through semi-structured interviews. Usability and evaluation trials were carried out at two of the partner sites (Sheffield Hallam University and Glenrothes College).

Typically a mixture of six academic and support staff were involved in the usability trials and evaluation interviews at each site; in addition interview scripts were completed by staff unable to attend on the day of the evaluation trials. The usability trials consisted of the users working through the TOIA system with an observer on the side making detailed notes on their activities, noting any problems or navigational issues which arose. The semi-structured interviews aimed to ascertain initial reactions to TOIA and to gain some understanding of how the tools might be used and what the associated barriers and enablers might be. The interviews are intended to help provide an impression of the culture within each institution and their respective attitudes to learning and teaching and, specifically, the implementation of assessment technologies.

Results

A preliminary analysis of the interviews has been carried out. A grounded-theory approach was adopted where themes emerge from the data. This section provides a summary of the main themes which emerged.

Training

The evaluation showed that most practitioners currently developing online assessments are primarily self taught. There is usually little institutional support and few formal workshops for them to attend. They get additional support through peers and by attending relevant conferences to find out about existing good practice.

Uses

Most of the online assessments being developed are being used for formative purposes, with summative assessment still being seen as high risk. There is a high percentage of the use of drill and practice type questions (particularly in the sciences and engineering), with some use of diagnostic and pre-course entry testing (for example in languages). There is an increasing interest in the development of automatic essay marking and mechanisms and tools for achieving this.

Barriers

Many of the barriers identified in the evaluation echo those found in previous studies (Bull, 1999; Warburton and Conole, 2004; Conole and Warburton, 2005). Practitioners lack the skills necessary to develop good pedagogically effective questions and complain of a lack of time to invest in this. Users are confused by the terminology used between different versions of the same online assessment tool and between systems. For example the terms 'test' and 'assessment' are often used interchangeably, similarly 'student' and 'candidate'. One of the general criticisms of the assessment tools available in virtual learning environments (such as WebCT and Blackboard) is that they lack a variety of question types, however it is worth noting that in reality many questions types are fundamentally the same, for example pull down menus or drag and drop questions can be considered as types of multiple choice questions. The important thing is to separate the pedagogical nature of the question from its presentational aspects, but it is also important to note that presentational aspects may have an impact pedagogically, for example drag and drop may have the same purpose and function as an MQC but because of its fundamentally different presentational aspects results in a different user experience.

In general understanding of CAA issues at institutional level is still poor and there is a lack of appropriate CAA-orientated policies and procedures in place in most institutions. Practitioners are dissuaded from investing the time in developing online assessment because there is a lack of personal recognition of their work. This is perceived as problematic in research-led institutions, particularly in the run up to the next Research Assessment Exercise (RAE). Users are concerned that the technologies are still not robust enough;

particularly for undertaking high-stakes summative assessments and that in general the ICT infrastructures in their institutions are unreliable. This is compounded by a feeling that there is a lack of appropriate technical expertise or experience in supporting online assessment, an issue which was highlighted in the Further Education sector.

Enablers

Personal interest and motivation was a key enabler, suggesting that much of the investment is still by the early adopters or e-learning pioneers. A major drive was the belief that initial investment in developing online assessments would be recouped in future years. A shift towards more of an institutional directive on virtual learning environments and online assessment tools in recent years was cited as important, with many institutions now committed to providing an institutional-wide site licence for an assessment tool.

Practitioners also reported that students' perceptions of online assessment were having an impact on whether they used CAA, many stating that their students had provided feedback saying that they liked CAA and considered CAA in some ways more 'objective' and more useful for their learning in terms of providing instant feedback and results. Where good central support was provided practitioners stated that this had helped in terms of increasing the confidence in and hence uptake of the use of CAA. Automatic marking, instant feedback and reusability of questions were all cited as important benefits of online assessment. Practitioners were becoming increasingly interested in using usage statistics in monitoring and tracking student learning. Online assessment, as with many other aspects of e-learning, was considered to help in making the process of teaching and learning more explicit and therefore perceived as having an important role in quality assurance processes. Practitioners valued the ability to learn from others by looking at sample question types that others had developed and then being able to adapt and repurpose these for their courses. Similarly there was a perceived benefit in terms of undertaking team developments, for example in the development of shared question banks. Two particular benefits were cited, the opportunity to learn from others and the ability to be able to provide a mechanism for peer validation of questions.

Issues

Just as a lack of personal recognition was identified as a barrier to investing time in developing online assessment, it was also clearly difficult to quantify the time allocated to developing online assessment and hence this made it difficult to take account of this in allocation of individual workloads. Despite an increased awareness of the importance of CAA at institutional level and the appointment of dedicated CAA officers in some institutions, in general CAA (and hence associated systems) is still not seen as mission critical and this is echoed in the fact that most institutions do not have a clear policy on CAA; neither are many set up to provide and support summative tests. Security and authentication pose major problems; there are usually no clear policies on submission procedures and these factors fuel the reluctance to scale up the use of CAA.

Finally practitioners worry about being locked into legacy systems, unable to transfer question and tests developed in one system to a new one, highlighting the importance of the development of Question and Test Interoperability (QTI) standards and interoperability more generally. The fact that TOIA is based on the QTI specification was seen as particularly advantageous.

Teaching strategies

Attitudes to the use of CAA are very much subject-dependent. Practitioners cited using a variety of teaching strategies in online assessment. Drill and practice was common, particularly in science and engineering subjects. Assessment was frequently used formatively throughout the course to aid student learning. Some teachers coupled this with the allocation of a small percentage of marks as a reward. Many practitioners used the date restriction facilities available in systems like TOIA to control access to online tests.

Testing was reported as being across a range of different types of skills and levels of learners, although as expected there was a weighting towards its use for first-year undergraduates and to test lower level skills. Practitioners adopted a variety of strategies for designing questions; a popular approach was to first write down the question as a multiple choice question, and then adapt and develop it into alternative forms. Many practitioners stated the importance of creating tests with a mixture of question types, and they felt strongly that there are some things that are not easily tested with multiple choice type questions. In this respect thought needed to be given to the meta-design of tests with an appropriate mix and sequence of question types. Overall they stated that devising questions types is a creative and iterative process which requires specialised skills and training. As assessment tools are becoming more sophisticated there is increasing interest in the use of adaptive testing and in particular the ability to link a series of tests together based on the student's prior performance.

Practitioners stated that they put a particular emphasis on designing tests which were interesting and which would motivate and engage students. One of the strategies adopted was the use of CAA as a means of assessing base level skills and using this diagnostically as a means of tailoring student workloads and assigning individuals with work at the appropriate level. Many teachers cited the use of multimedia as a valuable means of increasing the authenticity of materials. One specialised use of CAA was in supporting special learner needs.

A number of other strategies were offered. For example getting the students to complete a self-test after working through a topic or as a complement to a face-to-face seminar where they are asked to read articles, complete a short online comprehension test and then discuss their findings in a seminar. Obviously an important advantage was the use of online assessment to provide ongoing formative feedback to students. Others stated that they used a series of phased CAA tests to ensure students were attending on a regular basis – allocating small but cumulative marks as a reward. Some had used elements of summative testing part way through or at the end of a module, but

this was still relatively limited. Finally, online assessment was also being used with work placement students, for example with students returning after a work placement as a means of easing them back into academic study.

Student experience

The evaluation also looked at the reported student experience of the use of online assessment. Monitoring of usage statistics indicated that pattern of student usage varies across day and night, weekday and weekends; supporting the notion that students like the use of e-learning and in particular e-assessment to enable them to undertake flexible patterns of learning. In general students reported that they liked online assessments. Students were increasingly accustomed to using computers in all aspects of their lives, although there was a weighting towards the use of computers for playing games. There are clearly benefits to this; students are comfortable using computers, but often they lack the necessary e-literacy skills to make effective use to support their study (such as skills for finding, evaluating and using information, and for developing effective online communication and collaboration strategies). Both students and teachers reported that students were motivated by getting an automatic score for the work and in particular stressed the benefit of instant feedback particular where this was detailed and of relevance to their understanding.

Question types

Not surprisingly there is some correlation between the type of questions used and the subject, with more objective, and drill and practice type questions being used in the science and engineering subject areas. All reported the difficulty in creating good question types and the specialised skills needed to achieve this. Multiple choice questions were the most popular type and the one used most frequently. There appear to be differences in the types of question used across the FE and HE sectors and in particular between use for academic versus vocationally based courses, although due to the small scale nature of the evaluation it is difficult to provide any further details or quantification.

Interoperability

There was unanimous agreement about the importance of interoperability and a surprisingly high degree of awareness amongst practitioners of the term. In particular the issue of effectively linking legacy systems was cited and the issue of the current inaccuracy of much information held centrally within institutions resulting in the need for duplication of data at the local level. However, practitioners raised a concern that it was important to consider the educational interoperability as well as the technical aspects, an issue which to date has perhaps been under explored in current standards discussions. This is where current thinking in the development of learning design and associated standards will be important. Perhaps surprisingly ease of transfer was cited as more important to practitioners than wholesale course exporting. There was general concern over the investment of time put into the use of one system which may then be wasted if materials could not be exported to a new system. Practitioners stated that it was particularly important given users'

concerns over the rate of change of technology and about their investment of time in developing materials within a system which might soon be outdated.

Question banks

Most people could see the benefits and value of developing question banks and shared online resources. However a number of issues were raised, such as ownership, IPR and quality assurance. It was also stated that there was a potential conflict between sharing resources and the opportunity for commercial exploitation. Interestingly this seemed to be of particular concern in the FE sector. There was also a potential conflict noted between the development of generic questions and the ability to personally adapt questions. However practitioners could see the benefit of developing local departmental or institutional question banks which would minimise some of the concerns raised above. A perceived benefit was in developing shared materials in terms of the potential for staff development, support and peer validation.

TOIA specific evaluation issues

Evaluation of the TOIA system overall was very positive. Most users felt it was usable and comprehensive. There were problems identified with the interface in earlier versions which were rectified in subsequent versions. Practitioners highlighted the need for different author and administration views. In comparison with other existing online assessment tools, users felt that TOIA provided a comprehensive set of tools and functionality, with a better range of questions than assessment tools available in commercial VLEs. It was considered easier to use than many existing commercial products and importantly support staff felt that it would be easier to support institutionally. Most users were happy with the range of questions available, although some users did want specific question types and stated that lack of these might influence their decision to use the system.

Overall a range of benefits were cited about the products, including the fact that it was quite self-explanatory, easy to use, flexible, interoperable, provided customisable style sheets, was good from a support perspective, was free, and that it offered a complete system. Perceived disadvantages were the lack of large choice of question types and concern over the long term maintenance of free software/ the conflict between free tools and commercial version.

The feedback from the usability and evaluation trials was then used to improve the system before it was released so that all usability issues were addressed. TOIA has now been released and is available freely to all UK higher and further education institutions and at a fee for other institutions (http://www.toia.ac.uk).

Conclusion

This paper has reported on the evaluation of the TOIA online assessment system and shows how an understanding of the barriers and the enablers to the use of CAA, coupled with a knowledge of the roles and functions needed

to develop an online assessment system were used to develop a user-focused assessment tool, and in particular how close formative evaluation of the use of TOIA was used iteratively to improve the system. The evaluation has raised a number of critical questions about the use of online assessment, in particular:

- As new and more user-focused assessment tools are developed what impact might these have on the development of new forms assessment?
- What new forms of literacy are required for students and staff involved in online assessment?
- How can assessment tools be more effectively integrated with other e-learning systems?
- What is the impact of increased use of e-assessment in the role of e-learning strategy and policy?

These are the central questions which we need to research and address if we can achieve the desired goal of maximising ICT's potential to improve assessment. The paper attempted to illustrate how a strong evaluative approach to software design with a particular focus on end user needs has informed the development of an assessment system, TOIA, which it is hoped will provide an appropriate vehicle to enable practitioners to make innovative uses of online assessment to transform learning and teaching.

References

Beetham, H. (2004), 'Review: developing e-Learning Models for the JISC Practitioner Communities', available online at http://www.jisc.ac.uk/uploaded_documents/Reviewv2.1.doc

De Freitas and Mayes (2004), 'Review of e-learning theories, frameworks and models' commissioned review report as part of the JISC-funded e-pedagogy desk study on e-learning models.

Carroll and Appleton, (2001), 'Plagiarism – a good practice guide', JISC report, May 2001, available online from the JISC Plagiarism Advisory Service, http://online.northumbria.ac.uk/faculties/art/information_studies/Imri/Jiscpas/site/jiscpas.asp

Conole, G. and Dyke, M. (2004), 'What are the affordances of Information and Communication Technologies', *ALT-J*, 12.2, 113-124.

Conole, G. and Warburton, W. (2005), 'A review of computer-assisted assessment', *ALT-J,* 13.1, 19-33.

Jones, C. (1999), 'From the sage on the stage to what exactly? Description and the place of the moderator in co-operative and collaborative learning', *ALT-J*, 7(2), 27-36.

Sclater, N. (Ed) (2004), Final report for the Item Banks Infrastructure Study (IBIS), JISC-commissioned report

Sclater, N. and Howie, K. (2003), 'User requirements of the "ultimate online assessment engine", Computers and Education, Volume 40, No. 3, 285-386.

Warburton, W. and Conole, G. (2003), 'CAA in UK HEIs – the state of the art?', 7th International CAA 2003, Loughborough, July 2003