Enabling Reuse of CAA by Design

Andy Crofts,
Heriot-Watt University
Learning Technology Centre, Heriot-Watt University
Riccarton, Edinburgh, EH14 4AS
Fax: 0131 451 3283 Phone: 0131 451 3290

Abstract

It takes significant resources to write good questions and to develop quality assured CAA resources, so they should be maintainable and shared as much as possible across departments, universities and the whole higher education sector via mechanisms such as test banks. CAA content is often difficult to reuse because of issues such as proprietary formats, technical obsolescence and changes of computer platforms. Some of these barriers to reuse can be reduced though logical definition. At the core of a maintainable automated assessment system there must be logical structures for assessment content, which are independent of the system used.

Automatically marked assessments have been used by students over the last 3 years at Heriot-Watt University using WebTest, a tool developed for delivery of assessments to large numbers of students via the Web. WebTest was designed for efficiency for lecturer, student, test developer and administrator, while having a rich set of functionality. A logical structure is achieved by the use of XML for the definition of questions and tests. A wide bank of questions covering summative, formative and self-assessments has been developed and reused using our specification (DTD), which is independent of the WebTest engine and is being published as an open standard. The DTD enables diversity in question design, a high degree of randomization, the inclusion of mathematical display and a wide variety of question styles and answer judging mechanisms. It also does not restrict the user to web delivery, as a question is defined independently of the physical representation.

As other projects take a similar approach it is worth looking for common formats between systems and particularly at the opportunity to share questions. Standards bodies such as IMS (Instructional Management Systems) have a role to play to attempt to establish common standards for data and meta-data. If structures can be agreed across the education sector then economies of scale will justify the development of high quality shared assessment resources.

URL for WebTest: http://webtest.ltc.hw.ac.uk/

Introduction

Many computer assisted assessment (CAA) delivery systems are now available, some of which are standalone, while others are built into online learning environments. At Heriot-Watt University we have developed WebTest, a system for delivery of tests via the web. Assessments of various types for a wide range of subject areas have been created in WebTest by the collaboration of lecturers from each department with educational developers at the Learning Technology Centre. Our experience indicates that in the process of creating and delivering automated assessments the major investment is in the creation of good content, particularly the basic questions. Past experience with CAA indicates that the questions will generally have a much longer life than the system in which they are first implemented.

I suggest that if we are serious in increasing the use of CAA, then we need make it easy to develop quality CAA implementations by making it easy to find, reuse, re-purpose and adapt CAA content, while avoiding limiting systems to only simple types of answers such as multiple choice. There could be significant advantages in the sharing of some CAA resources between lecturers, departments and institutions. Reasons for sharing include financial savings, time savings and making the best use of scarce authoring and technical skills. There are many types of reuse of tests that should be considered:

- Reuse of the same test by a student. This could be for retakes of a test or in a self-assessment context.
- Use of the same test by many students in a class in both timetabled and open access contexts.
- Shared use of questions via question banks for subject areas.
- Reuse of test in a different software system to that in which it was developed.

There are many levels at which reuse could take place from single questions to a set of assessments for a whole module or study programme. I will mainly refer to questions, as these are the easiest level at which reuse could take place.

In considering reuse it is worth identifying what existing and potential barriers exist to the sharing of assessment content. Here I intend to focus on the technical aspects, which include:

- Most assessments need to be refined or adapted in some way for local conditions. Typical changes include variations in emphasis due to syllabus differences and lecturer preference, as well as presentational matters such as having the correct institutional logo and house style.
- In looking at many resources it is not easy to see the totality of the resource from a quick look at the visible content. Descriptions of content are needed. If content is to be shared then any assumptions need to be recorded.

- Different delivery systems currently have quite different ways of coding questions and other aspects of an assessment system such as formats for results.
- Lack of confidence in resources that are not produced locally. CAA is more mission critical than most computer based resources in that it has to work correctly and predictably every time. Ideally CAA resources would be checked for quality, both technically and educationally. This process can take longer than writing the questions.
- If assessments are shared between institutions then there may be problems with cheating and plagiarism. If standard assessments are used on different dates at different institutions, then the answers may be shared between students.
- The absence of standard interfaces to other parts of learning and administration systems. A key issue here is outputs such as results.

Creating CAA resources takes significant effort and time from some or all of lecturers, educational developers and system administrators. We find that lecturers are concerned that their investment in creating assessments is worthwhile in the long term and that their efforts are not lost due to changes imposed on them by factors such as:

- Changes in hardware and software. For example widespread use of the web in higher education was not anticipated by most of the early UK TLTP projects.
- Institutional decisions, such as standardization on particular software tools. For example a decision to use an online learning environment across a university may mean that content from an existing system needs to be transferred to the new environment.
- The need to deliver in different mediums. For example a distance learning course may create a need for stand alone self assessment web pages for a CD-ROM, instead of server controlled assessments which rely on an Internet connection.
- The need to make learning materials accessible to those with disabilities. For example a large print version of a test may be required.

In many ways those interested in content need to plan for reusability of assessment resources and even for legacy situations where content has to be extracted and taken to a new software system. It is essential to see the content of assessments as a valuable resource, which is likely to be more valuable than the system in which it is implemented.

The Heriot-Watt WebTest

WebTest is an automated assessment tool that is currently used in most departments at Heriot-Watt. Its main use has been in the efficient delivery of tests to large classes based on campus, but it has also been successfully used in distance learning courses. It is used for some class tests, but more typically students are allowed to take a test at a time and place of their choice. Randomization of question choice and of parameters in questions is used to give different, but equivalent instances of tests. This means that the same test can often be re-used by a student as a tool for the generation of examples for self-assessment and for credit in a continuous assessment situation. In these situations results can be given immediately and the immediate automatic feedback is then used by the student as the basis for further study. Examinations can be delivered by restricting delivery to specified dates and particular computers on the Internet.

WebTest differs from many assessment systems in the extensive use of randomization, though other academic projects have similar approaches. (Thelwall, 1998; Beevers, 1991). The randomization is the key to re-use, as it reduces the risks of plagiarism and cheating that would arise if students were all given the same test. It is possible to let students have instant access to their results and to the expected answers, while making the same test valid for continuous assessment.

We decided to create a web-delivered assessment system because existing software did not deal well with the judgement of numeric and algebraic answers and did not allow us to use randomized parameters in questions (Foster, 1997). The main attraction of the web was that it offered an interface that students seemed able to use without instruction. As well as simplifying and widening access to tests, by making delivery to the student platform, the web gave access to solutions for the display of many media and allowed for user customization, thus giving some support for disabilities.

In the design of WebTest we set out to create a powerful and adaptable system for the delivery of assessments in any subject. We could identify input mechanisms and judging algorithms for many types of answers, but also wanted to make the system extendable for future ways of entering and judging student input. Above all it seemed essential to make it easy to create content which could be easily adapted as the question is refined. For this reason we made a clear decision to separate the content and the engine that delivers and marks the tests.

We were attracted to the structure of HTML to code questions and decided to add extra tags to it to control the flow of a question. In doing this we realized that we were effectively re-inventing Standard Generalized Markup Language (SGML), an established standard for the logical representation of documents. Looking at established developments in SGML we confirmed that there were huge advantages to representing a test in a logical form and letting the browser take care of physical aspects such as text size and layout of images. A move towards SGML also gave us immediate access to standard tools for the editing of questions.

Since our first development a key development has been the creation of Extensible Markup Language XML, a simpler form of SGML developed as a data format for structured document interchange on the Web. Both XML and SGML have the concept of a Document Type Definition (DTD), which defines the rules for structuring a document.

A large bank of questions covering summative, formative and self-assessments has been developed using our DTD, called WebTest Markup Language (WTML). It structures our questions in a logical way which is efficient for delivery through the WebTest engine, but still enables diversity in question design, a high degree of randomization, the inclusion of mathematical display and a wide variety of question styles and answer judging mechanisms. The focus has been on ensuring that real questions can be asked in a natural way and still be automatically marked. Writing questions in a logical form should not limit the question design - as long as the structure used is well designed. Adherence to XML has made this goal easier to achieve.

WTML is independent of the WebTest engine and will be published as an open standard. It also does not restrict the user to web delivery, as a question is defined independently of the physical representation and can therefore be output in several forms.

Reuse of a Question in a Different Delivery System

In working with lecturers in the teaching of languages, my colleagues at the Learning Technology Centre initially used WebTest for self assessment tests, but this approach was felt to have disadvantages in some situations. A need was identified for more informal self-assessment tests, where marking would take place on the student's computer. A possible technical solution was through the use of dynamic HTML. Since they did not want to re-code questions in a different form, the standard WebTest authoring tools were used to produce questions in WTML. They then created a system to output a dynamic HTML page where each answer is individually marked, hints are available and answers can be revealed if the student is stuck. The production system used a standard XML processor and an XSL style sheet that defined the transformation from the logical WTML specification to the HTML output. Figures 1 and 2 show the two different outputs for the same question file. In figure 1 a standard web test is delivered to the browser from a server and the whole test has to be completed before marking takes place. In figure 2 each answer element can be marked immediately. Also a hint button and reveal buttons (displayed as a question mark) are provided for students who need extra help to get started.

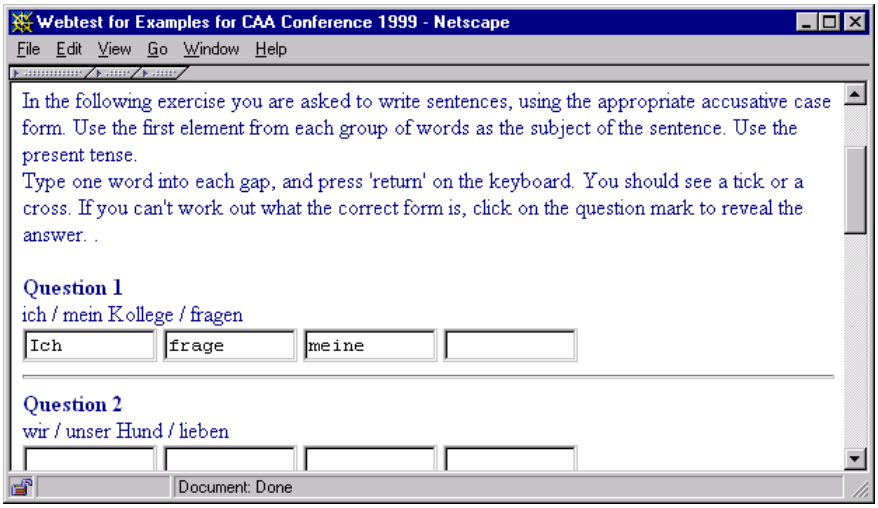

Figure 1. Delivery in a WebTest

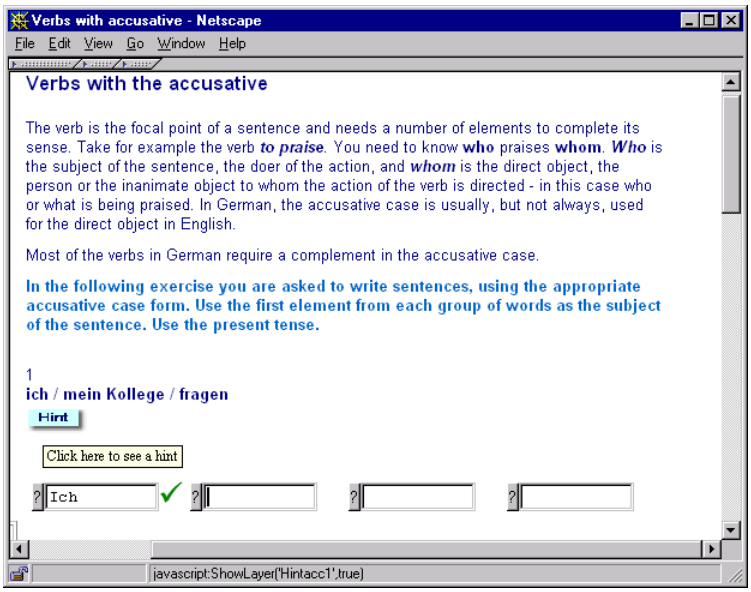

Figure 2. Delivery as dynamic HTML

The strategy of having multiple outputs means that decisions on usage, such as whether to use a question in a self-assessment test or in a submitted assignment can be made after it has been initially implemented. Thus all questions can be initially placed in a central question bank and then combined in a flexible way by the lecturer for each module. It would be possible to use the same technology to output tests in many other forms such printing to paper, thus facilitating further reuse.

Cataloguing Questions for Reuse

In looking at reuse of questions via question banks one issue is the need to classify, index and catalogue existing content, so that educators can search for relevant content. Metadata tagging provides a workable solution. The Instructional Management System project (IMS, 1999) is developing standards in this area for all educational resources, including questions and tests.

Tagging with metadata at the question level is a large commitment for the question author. We have found that, even with a very limited set of tags, which we are currently trialling in WebTest, there is often more metadata than question content. There has to be a powerful incentive for an author to tag questions with metadata and to maintain the metadata as the question is refined and used in various contexts.

Interoperability and Conversion

While interoperability of CAA systems sounds like a good solution, it is not a trivial matter to agree what CAA systems should share. One possible scenario would be for all CAA systems to conform to a common format for questions and tests. Indeed the development of such a specification is currently under discussion by the Instructional Management System project (IMS, 1999). The current proposed specification is based on two other markup languages QML (http://www.qmark.com) and TML (http://www.ilrt.bris.ac.uk/mru/netquest/tml/), both of which have been published as open standards. If a standard could be agreed then it would make it technically very easy to share content via mechanisms such as question banks.

Heriot-Watt, Strathclyde and Glasgow universities are collaborating on the Scottish Computer Assisted Assessment Network, a project funded by SHEFCE under their Web Tools Initiative. One strand of work under the project is to promote engine independent formats for questions, probably as an XML DTD.

There are significant problems with defining interoperability and in trying to make it cover all of the current practice in assessment. At the question level it would probably be straightforward to agree a common format for core question types, such as simple one part multiple-choice questions. In reality a rich set of question types is used in systems like WebTest or Clyde Virtual University (Whittington, 1998) and it is harder to see how formats could be agreed for interoperability of these questions. Translation between question formats may be a better route. Use of XML does not guarantee that

translation can take place, but its use does give a clear specification of question structure, and conversion to new DTD should be possible with the use of a style sheet.

Whatever directly emerges from current projects in looking at interoperability the standards under discussion should clarify many of the issues around reuse.

Conclusions

Reuse of CAA content can and should be designed into systems used for creation and storage of content, where some issues that prevent reuse are of a technical nature. CAA storage systems should:

- allow reuse of content in any conceivable form.
- allow content extraction for translation to other systems.
- use open standards where possible and avoid proprietary formats.

XML compliant systems seem the best current way to achieve these goals. Standards for interoperability of content between systems are under investigation and could increase reuse of CAA, though there are still many issues to overcome.

References

Beevers, C.E., Cherry, B.S.G. , Foster, M.G., McGuire, G.R. (1991), "Software Tools for Computer Aided Learning in Mathematics" Avebury Technical.

Foster, M.G. and Crofts,A.C. *WebTest and Mathematics Service Courses* CTI Maths and Stats Newsletter, November 1997.

Heriot-Watt WebTest http://webtest.ltc.hw.ac.uk/

IMS Project http://www.imsproject.org/

Thelwall, M., (1998) "The Advantages of Randomly Generated Computer Assisted Assessment", Proceedings of the Second Annual Computer Assisted Assessment Conference, 111-116.

Whittington, D., (1998) "There's more to the Web than Multi-Choice", Proceedings of the Second Annual Computer Assisted Assessment Conference, 117-123.