

INTEROPERABILITY WITH CAA: DOES IT WORK IN PRACTICE?

Niall Sclater, Boon Low and Niall Barr

Interoperability with CAA: Does it Work in Practice?

Niall Sclater
Boon Low
Niall Barr
University of Strathclyde

n.sclater@strath.ac.uk

Abstract

IMS has been promising question and test interoperability (QTI) for a number of years. Reported advantages of interoperability include the avoidance of “lock in” to one proprietary system, the ability to integrate systems from different vendors, and the facilitation of an exchange of questions and tests between institutions. The QTI specification, while not yet an international standard for the exchange of questions, tests and results, now appears to be stable enough for vendors to have developed systems which implement such an exchange in a fairly sophisticated way. The costs to software companies of implementing QTI “compliance” in their existing CAA systems, however, are high. Allowing users to move their data to other systems may not seem to make commercial sense either.

As awareness of the advantages of interoperability increases within education, software companies are realising that adding QTI import and export facilities to their products can be a selling point. A handful of vendors have signed up to the concept of interoperability and have taken part in the IMS QTI Working Group. Others state that their virtual learning environments or CAA systems are “conformant” with IMS QTI but do these assertions stand up when the packages are tested together? The CETIS Assessment Special Interest Group has been monitoring developments in this area for over a year and has carried out an analysis of tools which exploit the QTI specifications. This paper describes to what extent the tools genuinely interoperate and examines the likely benefits for users and future prospects for CAA interoperability.

Introduction

The main objective of the IMS QTI specification is to provide a standard format in which assessment information can be interoperable and reusable among different systems. The processes involved in CAA include question authoring, storage, administration and rendering. This study focuses on two of these areas, authoring and rendering, to discover if various assessment tools meet their claims to utilise the IMS specifications and actually allow assessment content to be exchanged.

There is a limited but growing range of supposedly IMS QTI “compliant” tools available. Five of these products are analysed in this study. They vary dramatically in the services which they offer and in the context in which they are intended to be used. These include:

- A leading commercial CAA package (*Questionmark Perception*)
- The two leading virtual learning environments (*WebCT* and *Blackboard*)
- A lesser known but highly innovative rendering tool (*Canvas Arena*)
- A university-built tool developed for testing IMS QTI (*CETIS Rendering Tool*)

Assessments can be authored with dedicated tools such as *Perception*. This is a fully-fledged CAA system, modular, and capable of addressing most aspects of CAA either as a standalone Windows-based solution or as a client-server web solution. Authoring in *Perception* is accomplished using two applications: *Question Manager* (for creating questions) and *Assessment Manager* (for retrieving questions from the databases and organising them in specific delivery contexts). Our evaluation focuses on *Question Manager*.

In addition to *Perception*, we evaluate the question authoring facilities of *WebCT* Campus Edition (v3.6), a web-based VLE server and *Blackboard CourseSite*, an online hosting service derived from *Blackboard 5*. Both of these VLEs have dedicated tools for creating and managing quizzes which can be integrated with their online courses.

The other two applications evaluated are rendering tools built specifically for delivering IMS QTI assessment. *Canvas Arena* is a new application developed by Sheffield-based company, Can Studios. It is part of the *Canvas Learning* integrated CAA design environment which also contains a forthcoming authoring tool (*Canvas Author*). *Arena* is based on *Macromedia Director* technology and facilitates content delivery in a number of ways including via web browsers using *Shockwave* and standalone executables for Mac and Windows platforms.

The *CETIS* (Centre for Educational Technology Interoperability Standards) *Rendering Tool* is an online demonstrator for testing the QTI specification. Developed by Graham Smith of Leeds University, it enables users to submit questions in QTI format to a server program which then renders the content and appropriate results feedback in HTML.

Scope of the Evaluation

Whilst the CAA applications selected for the evaluation consist of diverse functionality, this paper concentrates on the *exchange* of assessment data between applications, and does not provide a comprehensive review of the internal functions (e.g. editing features and user interface) of the applications. A brief evaluation of question creation in three of the applications is given. The capability of importing and exporting questions to and from the applications is also compared. The importing evaluation uses a suite of test questions based on two model test questions.

Test questions

Two model questions were defined for the purpose of the study; both are multiple-choice questions with four options. As one of the strengths of the QTI specification is the flexibility of its system of incorporating feedback this was given prominence in the question design; both questions have separate feedback defined for each possible answer. The first question consisted only of unformatted text and the second question added some formatting to the question (a bold section) and had images for the answer options. In addition to the individual feedback for each possible choice a further section of feedback consisting of both text and an image was defined for the incorrect answers. All images were put on a web server, and the full URLs were used to define the image locations in the tests.

The first question in our simple electronics quiz should be very easy for any system to render and process as it contains only unformatted text.

Question 1:

Which electronic component allows alternating current (AC) to flow but prevents direct current (DC) flowing?

- A transistor
- A resistor
- A capacitor *[correct]*
- An inductor

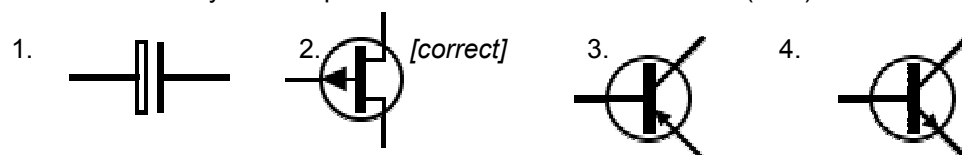
Feedback:

- Transistors are used to amplify or switch both AC and DC currents, the correct answer is capacitor.
- Resistors impede the flow of both AC and DC currents, the correct answer is capacitor.
- Yes, capacitors prevent DC current flowing.
- Inductors have high impedance for AC currents and low impedance for DC currents, the correct answer is capacitor.

The second question adds the simple features of a bold section in the text and images for both options and feedback. It also has the feedback for wrong answers split into two sections, one answer-specific and one generic.

Question 2:

Which of these symbols represents a **Field Effect Transistor** (FET)?

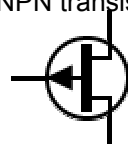


Feedback:

1. No, the symbol you selected represents an electrolytic capacitor.
2. Yes, you selected the correct symbol
3. No, the symbol you selected represents a PNP transistor.
4. No, the symbol you selected represents an NPN transistor.

Generic feedback for options 3 & 4:

The symbol for a Field Effect Transistor is:



Each question was converted to QTI 1.1 and 1.2 XML by hand and checked against the DTDs using TurboXML. The version 1.1 versions were also very slightly simplified – in the first question the feedback for incorrect answers was put into a single section, and in the second the extra piece of feedback was added to each incorrect answer's feedback individually. The 1.2 version of the second question was the only one that actually contained mark-up that would not be valid with the alternative DTD.

In addition to the four hand-coded test questions (the two model questions validated in two version of QTI), we replicated the model questions using the applications capable of authoring and exporting questions in IMS QTI specification: Perception Question Manager and WebCT. As a result, a suite of eight test questions was used in the evaluation. Further details of the test questions are given in the Appendix.

Criteria and Method

The following criteria were used to evaluate the applications. The criteria were selected specifically to test the question import and export functionality of each application. The criteria fall into three distinct categories:

Question creation – the ease with which the two model questions can be created and the degree of accuracy of question rendering within the following applications:

- Perception Question Manager
- WebCT
- Blackboard

Exporting – the ease and accuracy with which the following applications export questions into the IMS QTI specification :

- Perception Question Manager
- WebCT
- Blackboard

Importing – the degree with which *the suite of test questions* can be imported and how accurately the imported questions are rendered. Besides the authoring tools, the applications to be evaluated in this section also consist of the two rendering applications:

- Perception Question Manager
- WebCT
- Blackboard
- Canvas Arena
- CETIS Rendering

Results of the Evaluation

Question Authoring

Not surprisingly the dedicated assessment tool, Perception, offers the most powerful question input system. Two options are available for input, an easy to use wizard system or a powerful editor. The first question was created with the wizard without difficulty. The second question was also entered using the

wizard, however the HTML bold section was not automatically recognised and there was no obvious way of adding images. A simple 'work around' was used – images were put into the wizard using HTML tags, and all sections of the question were subsequently converted from 'text/plain' to 'text/html' using the editor.

The authoring tool of WebCT was easy to use, and allowed questions, answer options and feedback to be typed in either as plain text or as HTML. Plain text was used for question 1 and HTML for question 2, using tags to put in links to the images. The feedback for Q2 had to be reduced to a single section for each possible answer, simplifying the logic of our question, but making no difference to what the student would see.

In Blackboard CourseSites, the authoring tool was also easy to use and accepted HTML by default. As with the other tools images were typed into text input sections using HTML tags. Feedback was restricted to one section for the correct answer and one for the incorrect answers, so for both questions the incorrect feedback fields had to be merged, and students would see the feedback relating to all answers when they submitted an incorrect answer.

Exporting

Exporting questions into IMS QTI format in Perception was straightforward, and the output contained no errors when compared against the DTD. A visual examination of the XML revealed that the images were referenced using HTML links within <mattext> sections, reflecting the way they had been input.

In contrast, exporting of the quizzes generated in WebCT requires the use of a command line tool (Content Migration Utility) that has to be run on the server, so is only available to administrators. The tool exports a full course in IMS format, with the questions included in QTI format. The question file contained no reference to a DTD, but instead specified a name space within the <questioninterop> tag – good XML, but not within the QTI specification and so flagged as an error by TurboXML. Other than this the file checked against the QTI 1.1 DTD correctly. Visual examination of the file revealed a couple of other errors – the <mattext> fields for the answers did not have the texttype='text/html' attribute included, and < characters in HTML sections were replaced with < rather than the HTML being put into CDATA sections – as our tests did not include any < characters this problem was not critical.

Initially it appeared that exporting was going to be simple in the Blackboard system – questions had to be moved to a 'pool' from the test and then exported. A zip file containing the exported questions could then be downloaded. Unfortunately it turned out that this only contained an IMS content manifest in XML format. The questions were in a proprietary format.

Importing

The eight test questions were imported into the applications. The degree and accuracy to which the questions can be imported is varied. For instance, a question may be successfully imported into an application, but it may not be

rendered correctly, or the results responding may be incorrect. Hence, we distinguish three levels of question importing: authoring (in authoring tools and VLEs), rendering (in all tools) and responding (in all tools). In each category, we also specify three extents of importing: yes (Y), no (N) or partial success (P). The test results are summarised in Table 1.

		v.1.2		v.1.1		Perception		WebCT	
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Import (Authoring)	Perception	Y	Y	Y	Y	-	-	Y	Y
	Blackboard	N	N	N	N	N	N	N	N
	WebCT	N	N	N	N	N	N	-	-
Import (Rendering)	Perception	Y	P ²	Y	P ²	-	-	Y	P ³
	Canvas Arena	Y	P ⁶	Y	P ⁶	Y	P ⁶	Y	P ⁶
	CETIS	Y	P	N ⁸	N ⁸	Y	Y	Y	Y
Import (Responding)	Perception	Y	P ²	P ¹	P ¹	-	-	P ⁴	P ⁴
	Canvas Arena	Y	- ⁷	Y	- ⁷	Y	- ⁷	Y	P ⁵
	CETIS	Y	N	- ⁷	- ⁷	Y	Y	Y	Y

Table 1: Results of the evaluation

¹ didn't show all feedback correctly – only the first was shown;

² didn't render online graphics specified with an URI;

³ didn't render online graphics choice correctly because of a fault in WebCT, '<', '>' were translated to ascii codes;

⁴ given the wrong feedback;

⁵ didn't render online graphics, but provide correct feedback;

⁶ didn't render online graphics, but showed the (HTML) bold text correctly;

⁷ due to corresponding rendering problem, test cannot be carried out;

⁸ DTD of version 1.1 not available online

Importing the test questions into Perception Question Manager was easy since the application contains an option for importing IMS QTI files in its menus. In contrast, both of the VLEs were only able to import quizzes specified in proprietary formats and do not support the importing of IMS QTI questions. WebCT does have a command line tool, its *Content Migration Tool*, for importing courses. However, the tool can only be used by a systems administrator on the VLE server and assumes prior knowledge of how WebCT courses (including quizzes) and proprietary tools are specified and packaged in particular packages. This makes it very difficult to import individual QTI questions. A solution would be to export an empty WebCT course using the utility and make the expected course as a template for importing quizzes. It would still be impossible though for end users to import questions.

Since both MLEs failed to import the test questions, they were not included in the rendering and response tests. Overall, these are good for the remaining three applications albeit with some errors. The rendering of online graphics in general can be problematic; the older version of QTI (1.1) also gives rise to

errors in rendering and results reporting. In one case, the rendering of the second model question in Perception was accomplished incorrectly as in figure 2. It would be easy to jump to the conclusion that Perception was at fault however an examination of the XML exported by WebCT reveals the problem. The image is enclosed in a test section that has no type specifier so Perception correctly interprets it as the default "text/plain" rather than "text/html".

The XML for the response option presentation exported from WebCT is -

```
<material>
  <mattext>&lt;img src="http://ford.ces.strath.ac.uk/images/ecap.gif"></mattext>
</material>
```

however it should be -

```
<material>
  <mattext type="text/html"><![CDATA[]]></mattext>
</material>
```

or more correctly, defining an image URL rather than HTML containing an image -

```
<material>
  <matimage imagtype = "image/gif" uri = "http://ford.ces.strath.ac.uk/images/cap.gif"/>
</material>
```

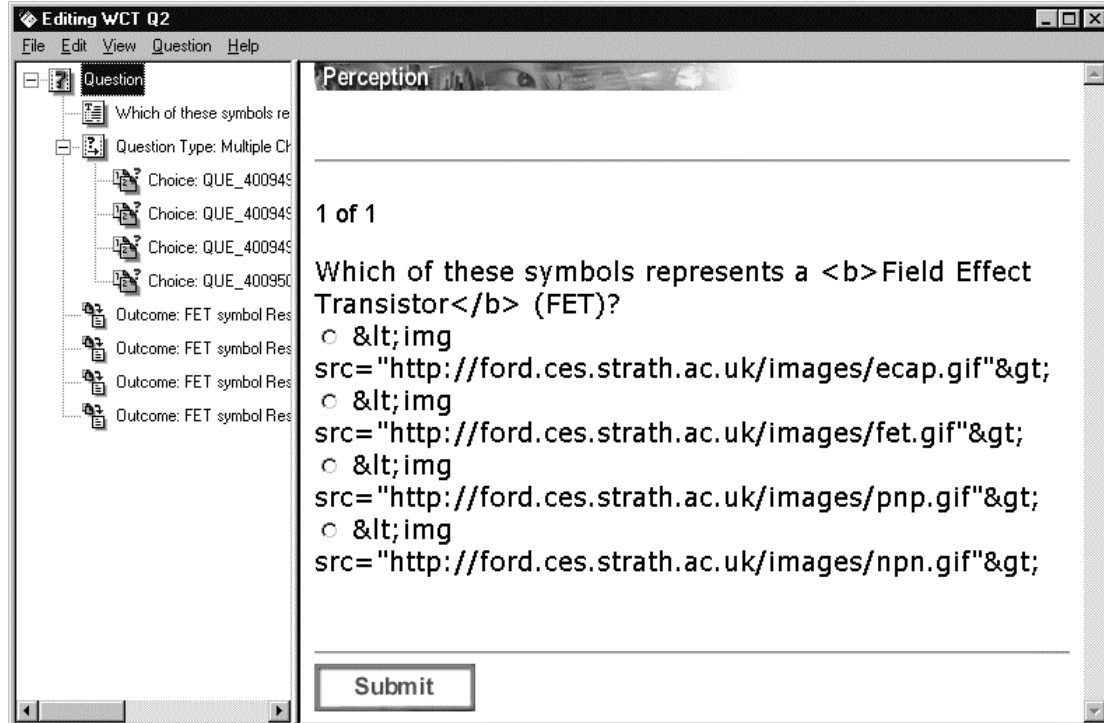


Figure 1: The transfer of Question 2 from WebCT to Perception

Conclusions

Our brief evaluation shows that it is now possible to transfer simple questions using an internationally agreed format among a variety of assessment systems. There are however a number of serious causes for concern. First of all, the VLEs tested proved very disappointing. Blackboard, which is a partner in IMS and has signed up to the philosophy of interoperability, failed the CAA interoperability test completely. Exporting your questions as an IMS content package but keeping the questions in a proprietary format does seem to defeat the object. WebCT has made some effort to implement the specification but the exporting facility can only be used by a systems administrator and produces code with errors. Importing was not possible with either VLE.

None of the other systems had problems in importing and rendering Question 1 which was, admittedly, one of the simplest imaginable CAA questions. Question 2 with its slight added complexity proved more problematic. QuestionMark, a major driver in the development of the QTI specification, has been more rigorous in implementing it than the VLE vendors. However even Perception failed to render graphics hosted on the web (a facility required by the specification) or to show all feedback correctly. Canvas also had problems with online graphics. Graham Smith's CETIS tool only partially rendered Question 2.

The expert knowledge of the specification and personal commitment which has gone into the creation of the three non-VLE products is known to be considerable. If, despite the programmers' expertise and the relative simplicity of the questions chosen, they are still failing to be rendered correctly does this bode well for interoperability? There is no doubt that the QTI specification is highly complex, with some remaining apparent inconsistencies and ambiguity making it difficult to implement.

Fortunately we are still in the early days and the problems that were discovered with these particular questions should be trivial to correct; it is likely that developers could iron out many such incompatibilities between their systems with ease. There has not yet been, to our knowledge, a QTI plugfest but it is perhaps time for such an event to take place. In addition to increased cross-vendor collaboration, stress-testing programmes which cover all the possibilities of QTI need to be devised and carried out by independent evaluators. These should include complicated layouts and feedback for different combinations of responses as well as incorporating the many other QTI response types.

More fundamentally, our study has not looked at how the engines deal with assessments and sections as opposed to single items (questions). Perception currently imports and exports only at the item level, a major drawback if you have a database of assessments marked up in QTI to transfer. Evaluations of another major area of the QTI specification, response processing, also need to be carried out.

While other parts of the two VLE packages examined may “conform” to other IMS specifications, their vendors’ efforts in the realm of CAA are not encouraging. Adding real interoperability to your product is neither cheap nor easy (as demonstrated by the problems encountered with Question 2). The main reason for vendors to do so is to give them a competitive edge. While it is helpful to be able to claim that your product is interoperable it is not necessarily to your advantage as a vendor for it to be so. As well as adding to your system development costs your clients may ultimately decide to move to another system and use your interoperability feature to take their content with them.

It may be that both Blackboard and WebCT are content to allow those who are interested in question and test interoperability to use QuestionMark (which effectively plugs in to their products anyway) and let it take the development cost hit. But for the many thousands of users who are developing banks of questions with these VLEs alone, they will find them impossible or very difficult to transfer out of these systems for the time being. Meanwhile other companies are moving into the arena. EQL, a competitor to QuestionMark, is hastily adding QTI facilities to its *I-Assess* system, having identified the marketing advantages to be gained from having an interoperable assessment system. We look forward to being able to test the interoperability credentials of such systems against the others in the near future.

Appendix

Test questions

Hand-coded

Q1: Multiple-choice question with multiple feedback - QTI version 1.2

Q2: Multiple-choice question with multiple feedback with graphics - QTI version 1.2

Q3: Same as Q1 with simplified feedback in QTI version 1.1

Q4: Same as Q2 with simplified feedback in QTI version 1.1

From authoring tools

Q5: Same as Q1, created in and exported from Perception Question Manager

Q6: Same as Q2, created in and exported from Perception Question Manager

Q7: Same as Q1, created in and exported from WebCT

Q8: Same as Q2, created in and exported from WebCT

These questions are available online at: <http://www.cetis.ac.uk/assessment/>

References

Smythe, C. and Roberts, P. (2000) *An Overview of the IMS Question and Test Interoperability Specification*. Proc 4th International Computer Assisted Assessment Conference, 21-22 June 2000, Loughborough