# QUESTION AND TEST INTEROPERABILITY: INTRODUCING VERSION 2 OF THE IMS QTI SPECIFICATION

**Steve Lay**

# Question and Test Interoperability: introducing version 2 of the IMS QTI specification

Steve Lay,
University of Cambridge Local Examinations Syndicate
IMS QTI Project Team Lead
s.w.lay@ucles-red.cam.ac.uk

## Introduction

The Question and Test Interoperability (QTI) specification describes an XML-based technical format for the coding and exchange of assessment content from individual questions through to complete tests. The specification was first published by the IMS Global Learning Consortium in June 2000 and since then an increasing number of developers have used it as a guide to implementing assessment functionality in their e-learning systems, in addition to the more established assessment tool providers who have added QTI export/import facilities to their products.

In September 2003, IMS approved a project to develop version 2 of the specification, concentrating on the format of individual 'items' and the way they can be packaged and integrated into authored learning experiences based on the companion specifications known as *Simple Sequencing* and *Learning Design*. At the time of writing, a draft of version 2 is being prepared for public review and should be available for review from the IMS website two weeks before the CAA 2004 Conference.

Version 2 will represent a significant step forward for the specification, addressing many of the issues that have been raised by developers and providing some powerful new capabilities based on some of the extensions to version 1 that have developed within the community. Readers are encouraged to review the public draft for detailed technical information. This paper will concentrate on the more general issues raised with version 1 of the specification and how they are being addressed. It also attempts to address some more general criticisms of technical standardization work as a whole.

## Why Standardize?

QTI is primarily a content format for assessments and their component parts. The desire to create a specification to describe assessment content has its origins in the companies that supply tools for carrying out Computer Based Assessment (CBA) and their customers. Content owners see a lot of value tied up in the assessment content they create and therefore seek to minimize the risks to that content by encouraging tool suppliers to publish information about the format their content is stored in. For example, Question Mark

Computing published their QML (Question Markup Language) specification to enable assessment content created with their tools to be used by others.

Although most of the early users of CBA were using systems that included support for authoring right through to delivery, larger content providers typically use a combination of tools from different suppliers to author, manage and deliver their assessments. The market for these tools (and services based on them) was segmented by the technical format used to pass data between them.

There are many reasons why standardization might been seen as beneficial. Certainly there is a pressure from content providers to encourage uniformity of technical formats within the market as publishing into multiple incompatible formats simply increases costs. Left alone, a *de facto* standard would probably have emerged but a proprietary solution presents significant risks for the organizations with an infrastructure built on one of the alternative technical formats.

One of the lessons learned from the often cited video tape standardization battle of the 1970s/1980s was that opening up a technical format provides a significant advantage to adoption. Another factor was the use of lower licensing fees to encourage partner companies to create complementary products and services. In the world of the Internet, and particularly within education there is an expectation that data formats will have no licensing fees at all, further encouraging organizations to pool their resources when developing them.

CBA has also provided an opportunity for embedding assessment into the learning process. This blurring of the distinction between summative and formative assessment suits the development of QTI in a broader e-Learning context. The IMS Global Learning Consortium provides this context and was therefore a suitable platform for companies and organizations involved in assessment to work together towards these shared aims.

## The Shortcomings of Version 1.x

Version 1 of the QTI specification took a huge step forward in the development a shared technical format for assessment content. It has also been a great success from the point of view of positioning itself as the leading technical format for assessment content internationally. The specification documents remain extremely popular downloads with visitors to the IMS website.

Even though QTI has been able to position itself as the dominant technical format for assessment content, how well does it achieve the aims of the end user? Is data stored in QTI format less risky than other formats? Will it be easy to translate QTI data into future technical formats and can it be transferred between different systems now?

In June 2003, Pierre Gorissen carried out *A Quickscan study on the usability of QTI for De Digitale Universiteit in the Netherlands*. The conclusion was that "it is possible to construct a basic set of QTI questions that can be imported by all the applications that support QTI in one way or the other". The report concentrated on 6 tools being considered for their project, all with QTI reading capabilities and has been extended to include a 7th with a more recent addendum. However, the report did emphasize that interoperability broke down when test data became more complex.

Many of the problems with more complex QTI files have been discussed by developers and solutions proposed through initiatives such as the QTI Common Practice Initiative coordinated by CETIS through the Assessment Special Interest Group based in Strathclyde. Some of these recommendations from developers have been incorporated into version 1.2.1 (published in February 2003) but many of the issues suggested changes that could not be made without breaking backward compatibility.

The popularity of QTI amongst tool developers demonstrates that they and their customers see benefits in the adoption of an open standard for assessment content. However, if QTI is to continue to be a success and progress to the level of "standard" it must deliver interoperability beyond basic examples and must better support the making and testing of compliance claims.

## An Overview of Version 2

Version 2 of the QTI specification is a complete re-write, but to make the task more manageable the scope has been narrowed to include just the individual items of an assessment and not the complex ways in which these items are combined to form complete tests. Narrowing the aims of the project sets clearer boundaries for what is in and out of scope while still supporting almost all existing implementations - the more complex examples of aggregated assessment content being a relatively recent addition to version 1 and only sparsely adopted to date.

Version 1 contained many alternative ways of achieving the same result. Example files that had evolved over the three years of development failed to give a clear indication of how the specification was to be used. Version 2 is more prescriptive. It has been developed in a much shorter timescale and this has enabled a more consistent set of examples to be developed. The aim of the specification has been to produce one mechanism to achieve each desired result.

Although backward compatibility has had to be sacrificed to achieve these benefits a clear migration path has been described. It is possible to automatically convert items in version 1 format to version 2 and this has been demonstrated by the inclusion of an experimental script in the specification package.

The following sections describe the main technical changes to the specification.

*Rendering Issues*

There has been a complete reworking of the material model to improve the division between structural information and presentation-style. It is acknowledged that HTML now provides the most widely adopted method of marking up structured text and a profile of XHTML has now been incorporated directly into QTI. This should also help systems provide renderings for a wider range of access requirements.

*Response Processing and Feedback*

Response Processing is the name given to the transformation of a candidate's responses into assessment outcomes (i.e. scoring). A new set of standard response processing templates has been provided that allow for a much simpler representation of simple use cases. In particular, it is now trivial to parse the correct response for a simple multiple choice item directly from the QTI file without having to evaluate any scoring rules. The feedback model has also been clarified and improved to enable much better support for formative assessment tasks.

*New Question Types*

A number of commonly used item types were simply not representable in version 1 without resorting to extensions (which are not interoperable). Version 2 has documented these item types directly. Overall there is an improved interaction model that clearly describes the behavioural requirements on delivery systems for each interaction (item) type. This is an important step towards more successful interoperability and compliance testing.

*Harmonization With Other Specifications*

Version 2 describes a method for packaging assessment content and media files for transfer between systems using the IMS content packaging specification. Not only does this change enable more reliable exchange of assessment content itself but it paves the way for sequencing rules to be added, for example to facilitate organizing items into assessments. Furthermore, the use of content packaging goes some way towards enabling formative assessment tasks to be integrated into overarching learning experiences. A suggested method of using assessment items in systems that support IMS Learning Design is documented and further recommendations on the use of IMS Simple Sequencing (and SCORM) are in development.

## Conclusion

By narrowing the problem domain, simplifying the data model, removing ambiguities and improving the documentation and examples, version 2 of the QTI specification represents another big step toward the goal of achieving a common standard for assessment content.

QTI tries to strike a balance between flexibility to encourage innovation and the needs of developers to have a specification that can be implemented and for which compliance can be sensibly tested.

The open nature of the specification has encouraged wide adoption across the education and training communities. For example, in the UK it has been recommended for consideration by the Office of the E-Envoy/DfES e-Learning Working Groups as part of the e-GIF.

## References


e-Government Interoperability Framework,
Version 6.0 April 2004
<http://www.govtalk.gov.uk/>

IMS Question and Test Interoperability Specification
IMS Global Learning Consortium
<http://www.imsglobal.org/question/index.cfm>

IMS Content Packaging Specification
IMS Global Learning Consortium
<http://www.imsglobal.org/content/packaging/index.cfm>

IMS Learning Design Specification
IMS Global Learning Consortium
< http://www.imsglobal.org/learningdesign/index.cfm>

An Overview of the IMS Question & Test Interoperability Specification
Colin Smythe, CAA 2000
<http://www.lboro.ac.uk/service/fli/flicaa/conf2000/pdfs/smythec.pdf>

Quickscan QTI; Usability study of QTI for De Digitale Universiteit
Pierre Gorissen, June 2003
<http://www.digiuni.nl/digiuni//download/EA183322-C145-0A52-7B7EF8A4EDFF4655.pdf>

Addendum to QTI Quickscan
Pierre Gorissen, November 2003
<http://e-learning.surf.nl/six/english/2078>

Question Markup Language (QML)
QuestionMark Perception
<http://www.questionmark.com/uk/qml/>

QTI Common Practice Initiative
CETIS Assessment SIG
<http://ford.ces.strath.ac.uk/QTI/cutiepie_index.html>