# Current Issues With Assessment Formats and Interoperability

Abelardo Pardo, Álvaro Agea, Carlos Delgado Kloos
Department of Telematic Engineering, Carlos III University of Madrid, E-28911 Leganés (SPAIN)

*Abstract*—**Assessment has been one of the areas in computer supported learning where technology has been quickly deployed. Support for computer based assessment is found not only in all Learning Management Systems, but also as stand-alone tools. This large number of tools has led to the appearance of a large number of formats to store, retrieve and exchange assessment material. Although institutions such as the IMS Global Consortium have proposed specifications (e.g. IMS QTI) aimed to facilitate the exchange of this material, in the actual landscape, there are still too many formats that significantly hinder the reuse of this material. In this paper an analysis of the implementation of these formats is described.**

***Keywords-assessment format, interoperability, IMS QTI***

## I. INTRODUCTION

In today's e-learning, the advance and support offered by technology is very uneven. Some aspects of a learning experience are fully supported by technology (for example assessment), whereas in others, technology is barely beginning to appear (for example, integration of generic services in a learning environment). Assessment has been one of the aspects where technology has had a very strong presence from the early times. Some computer-based assessment tools even pre-date the appearance of some learning management systems (henceforth simply LMS).

The issue of storing assessment material is intrinsic to computer supported assessment. Once a scenario where any form of computer supported assessment is performed the problem of how the material is stored and retrieved already appears. There are two dimensions in which this problem can be solved. The first one is considering only a single tool or product. In other words, given a tool for computer-based assessment (or an equivalent functionality present in a LMS), how can the assessment material be stored and retrieved by any instance of this tool.

But this issue, combined with the wide variety of solutions for computer-based assessment that are currently available, poses a second problem, that of interoperability. Suppose a scenario where Institution A has a rich set of assessment material that created using Tool A (or its equivalent service in the corresponding LMS). Analogously, Institution B has also a rich set of assessment material, but it was created with Tool B. Both tools are totally unrelated and therefore, store the assessment material in completely different formats. So far, none of the tools has acknowledge the other format as one that is supported. These two institutions have identified the potential of sharing this material, but it is not possible with the current formats.

The problem, then is to go beyond formats locally used by specific tools to a format that allows the exchange of assessment material among unrelated tools. The approach followed by international bodies such as the IMS Global Learning Consortium [1] has been to gather a set of experts in the field, analyze the current scenarios where computer-based assessment was being used and propose a format such that any tool would be using to both export and import assessment material. This format proposed by the consortium is IMS Question and test interoperability [2], henceforth IMS-QTI.

The appearance of this specification changed the landscape of assessment formats. For any tool providing computer-based support, the question of supporting IMS-QTI appeared. But the evolution of this specification over the last nine years shows numerous examples of the effect of several design decisions. Of course, it easy now to evaluate how appropriate a specification is to represent assessments. But computer based assessment has reached a stage which can be considered main stream and therefore there is some value in looking back and see how the once foreseen effects really shaped. In the following sections an analysis of the current landscape with respect to this format is analyzed. First a brief description of the aspects covered by IMS-QTI and the different available versions is given. Then a brief account of the type of support offered by the different tools is described. It follows the description of a case study in a scenario in which true cross-platform interoperability was needed. Finally, some conclusions are proposed to improve the level of interoperability among current computer-supported assessment tools.

## II. IMS QUESTION AND TEST INTEROPERABILITY

The IMS Question and Test Interoperability format was initially proposed by the IMS Global Learning Consortium and its version 1.0 was released as a public draft in February of 2000 and as a final specification in May of the same year. The proposal was based in QML, a structured language proposed by Question Mark Computing Ltd [3] in 1997. The idea behind this format is to capture the structure of assessment material with as least information as possible about how it should be visualized. The format used a syntax similar to an XML document, and even an XML Schema definition has been published.

978-1-4244-6571-2/10/$26.00 ©2010 IEEE April 14-16, 2010, Madrid, SPAIN
IEEE EDUCON Education Engineering 2010 – The Future of Global Learning Engineering Education


*Figure 1: Evolution of the Question and Test Interoperability Specification*

In general, IMS-QTI describes the structure for representing two elements: the question, and the assessment. The structure of a question includes information about the body of a question, possible answers, feedback and even grading procedures. The structure of assessment is more related to how a set of questions are organized, subdivided into sections, and how are the values obtained from individually grading the questions are combined to obtain an assessment grade. With this structure, the objective of this format was to allow the exchange of three elements of computer-based assessments: items (that is questions), assessments (different groups of items) and results.

After the first version, the consortium published version 1.2 two years later in 2002 which included a much more comprehensive set of documents describing its structure. Aside from a description of the information model, XML bindings and a best practice guide, version 1.2 also included documents describing how to select and order items, how to process outcomes, how to report results and an overview document. The evolution of the specification is shown in Figure 1.

But the aim to cover the three aspects of assessment management (questions, assessments and result manipulation), led to a specification that was fairly elaborated. As a consequence, a simplified variation of QTI version 1.2 was released with the name of QTILite. QTILite is a subset of QTI version 1.2 that represents an entry level into version 1.2. The main simplifications were:

- Reduced number of question types: Yes/No, True/False, Likert scales and multiple choice.
- Simple response processing with default mechanisms.
- No support for hints, solutions, etc.

As a consequence, tools supporting QTI had to choose between two versions (one a subset of the other) to guarantee some degree of interoperability. Some small amendments were introduced in Version 1.2 to obtain version 1.2.1 in March of 2003.

The next significant leap in the specification came with version 2.0. The assessment process needed to be taken into account but as yet another piece of the complex mosaic of e-learning procedures. For example, an assessment could be included in a more complex set of activities that need to be sequenced following certain rules. The result of an assessment could be used to decide the sequencing of activities (as it is conventionally done in multiple learning scenarios). Version 2.0 was conceived to accommodate this new emerging reality as well as to polish certain aspects from version 1.2.1. This major version change came when several commercial e-learning systems had already included support for QTI v 1.2.1 and meant an additional effort to accommodate the proposed changes.

But to guarantee that this reviewing process reached the e-learning community quickly, the scope of the changes was intentionally reduced to individual assessment items. The aspects of aggregation of items into sections and assessments and result reporting were left out of the review.

These two aspects were reviewed and published as a first public draft (revision 2) a year later in June 2006. Both aggregation and result reporting support were reviewed. At this point, tools supporting version 2.0 of the specification were scarce, and the lack of activity in this draft to become a final specification led the consortium to remove it. The reason was the lack of activity to push this specification forward. As a consequence of this decision, numerous institutions complained that they needed the specification to guide their current developments. As a conclusion, it seems that there is a number of companies and institutions using version 2.1 of QTI, but the participation of these institutions in the process to push the specification to a final stage is scarce. Although the specification was later re-published, the consortium warns that as is, the document is incomplete and in the process of evolving it based on input from the market participants.

| | | |
|---|---|---|
| Angel | | 2.1 |
| ATutor | 1.2 | 1.2,<br>2.1 (experimental) |
| Clix | 1.2 | 1.2 |
| DB Primary | | 2.0 |
| Diploma | 1.2, Lite | 1.2, 2.1, Lite |
| Dokeos | 1.2 | 1.2, 2.0 |
| .LRN | 1.2 | 1.2 |
| Moodle | 1.2 | 2.0 |
| OLAT | 1.2 | 1.2 |
| QTI Tools | 2.1 | 2.1 |
| QuestionMark Perception | 1.2 | 1.2 |
| Respondus | | 1.2 |
| Sakai | 1.2 | 1.2 |

*Table 1: QTI support in different tools*

The future of the QTI specification is a bit uncertain. Although version 1.2.1 seems to be widely used (the IMS Common Cartridge initiative [5]includes it as part of its formats), and the need for an improved version is perceived, there is not a clear path on how to obtain such improved version.

## III. TOOLS SUPPORTING QTI

With this landscape, tools for computer-based assessment started to choose which version of the QTI specification support. A summary of this support in only a subset of e-learning tools is shown in Table 1.As it can be seen, version 1.2 achieved a fairly high level of support. This support has been increased as a consequence of this version being in included as part of other initiatives within the consortium such as IMS Common Cartridge [5]. This format bundled together two previously created specifications, IMS Content Packaging version 1.2 and IMS QTI version 1.2.1 together with additional meta-data to manage learning content from the point of view of the producers as well as the users.

The final consequence of this situation is that when it comes to capture the structure of assessment material such that it can be exchanged among different tools, the most widely supported format is version 1.2.1 which has not evolved since 2003.

## IV. ASSESSMENT INTEROPERABILITY: A CASE STUDY

In order to experience first hand the problems that arise when trying to exchange assessment information among unrelated LMSs, an case study was prepared. The study has been carried out within the frame of the ICOPER project. This project is a Best Practice Network that started in September 2008, funded by eContentPlus programme of the European Commission. As part of the ICOPER objectives, an analysis of the best practices in the scope of assessment, and more precisely in the issue of interoperability has been proposed. The project must explore also the different level of adoption of the specifications and standard currently used in e-learning, and the issue of interoperability has been identified as one of the main burdens to overcome to increase adoption.

The case study was designed to include several institutions related to higher education (both educational institutions and commercial vendors) and explore the problems that appear when trying to exchange assessment material. The organizations involved in the study were the Open University of the Netherlands, Giunti Labs, IMC Ag and Carlos III University of Madrid.

The objectives of the study were:

- Obtain a sample of the tools and formats being used to perform computer-based assessments
- Identify the interoperability problems

The type of interoperability problems that were specifically targeted were:

- Exchange of assessment material at the level of "exams", therefore, not only at the level of question items.
- Exchange of generic exam annotations.
- Connection with Learning Outcomes.

The LMSs used by the institutions participating in the study is shown in Table 2.

| | Moodle | .LRN | Clix |
|---|---|---|---|
| UC3M | X | X | |
| Giunti Labs | X | | |
| OUNL | X | | |
| IMC | | | X |

*Table 2: LMSs used by the different institutions in the study*

The first step was to analyze the import/export capabilities of the LMSs to see which information exchange flows were possible. The format that was commonly supported by the three sampled LMSs was (as expected) IMS QTI. The support for QTI import/export in its different versions is shown in Table 3.

The first observation from the table is the total absence of any kind of support for QTI version 2.1. These three LMSs were implemented when version 1.2 was available, and the migration to the new format was not considered. A different situation appears when considering QTI 2.0. A partially implemented export module was found but with significant shortcomings. The module simply placed a subset of the information contained in the questions in certain special fields of the QTI file but certain special information was not included. The file exported by Moodle then, was not possible to be used in any other platform (nor even Moodle itself because it lacked the import functionality).

| | QTI 1.2 | QTI 2.0 | QTI 2.1 |
|---|---|---|---|
| Moodle | No | Export | No |
| Clix | Import/Export | Import/Export in development | No |
| .LRN | Import/Export | No | No |

Table 3: Import/export support of different QTI versions

Several attempts at providing support for QTI in Moodle were detected in the developers community, but none of them was considered mature enough to be considered. A special situation was seen with respect to the Respondus [7] tool. Respondus is a tool for designing and deploying assessment in e-learning platforms. In principle the product is self-contained, that is, it offers an authoring environment to produce assessment and a deployment environment to use them in a learning experience.

But in order to increase the interoperability, the company produced what is called a "Respondus plug-in" for Moodle that imports assessment material create with these products into a Moodle platform using QTI version 1.2. Being a separated assessment product, the company already has modules to incorporate assessment material created with their product suite into the main LMSs (Atutor, Blackboard, eCollege, etc.) but in the case of Moodle, and because the product is open-source, the uploading of material is based in QTI version 1.2.

This plug-in could be considered as a suitable vehicle to import generic QTI documents into Moodle, but a closer inspection quickly revealed that the QTI produced by

Respondus to be then imported into Moodle was not compatible with the other LMS in the study. The company itself clarifies that the module is not a regular QTI importer but one specifically for the Respondus system. The main reason to avoid claiming to have a generic QTI version 1.2 importer is because only certain uses of the specification are supported. A generic QTI file using other aspects of the specification cannot be handled by this importer.

As with the rest of LMSs, in principle, a potentially effortless exchange of material was possible between .LRN and Clix (if QTI version 1.2 was used). However, upon a closer inspection, again the difference in implementation in the import/export functionality of different platforms made this possibility non-trivial.

The main difficulty when exchanging assessment material is the level of flexibility offered by the specification. There are several forms in which a set of questions making an assessment can be included in a file and comply with the QTI specification. The existence of these different versions, although potentially positive from the point of view of flexibility it is a huge obstacle for interoperability. The amount of possible scenarios to consider when writing an import module derived from the specification rules it too complex.

Furthermore, the specification, to increase flexibility, leaves a large number of elements totally optional. This functionality also poses problems. For example, institution A exports a large number of true/false questions, but none of the questions in the file mark the correct answer. The file would still be in compliance with QTI, yet, most of the current tools would rule it unusable, or even worse, produce an incorrect import.

The trade-off between flexibility and widespread use needs to be carefully considered. Widespread use is very important for a specification such as QTI to be truly useful. But if the adoption is in part fostered by a specification with too much flexibility in its definition, the adoption will be uneven and interoperability will be significantly reduced.

A promising venue that is being explored to compensate this situation is the use of "profiles". A profile is a simplified or reduced version of the specification derived from a concrete scenario of application. Several initiatives have appeared that have contributed different QTI profiles such as, for example, questions in the area of mathematics.

The interoperability problem can then be re-stated in terms of profiles. Two platforms may exchange assessment related information effortless as long as they are encoded using the same QTI profile. But again, the essence of a common specification is lost.

Together with the problems when exchanging information about assessment, the study case revealed some interesting new venues to explore. None of the institutions participating in the study was annotating assessments or question items with information obtained from their use.

Although QTI, with its great flexibility, allows the inclusion of almost arbitrary meta-data, this type of annotations, although perceived by all participants as relevant, were not present in any of the used samples. The process to back-annotate assessments is still far from main stream and requires still an intuitive tool to facilitate the work. This type of annotation becomes very important when assessment material is shared in a wider scope such as federated searches of learning material.

There are currently multiple initiatives to offer federated content available to multiple institutions. The key feature to offer is a powerful search functionality. But a powerful search itself is based in a powerful labeling of the pool of available objects.

|In the context of assessment, if a teacher is searching for relevant assessment material, the search engine should know what type of questions are included in an assessment, if they have been used successfully, the type of scenario in which they were used, etc.

An the third shortcoming found in the study is the lack of connection between QTI and the higher level instances in which it develops. An assessment is usually tied to some sort of learning unit. Even worse, the objectives of a learning unit (in the most general sense) should be connected with the assessment methods, and more precisely with assessments described in QTI.

But this connection between assessment and learning outcomes (learning results, learning objectives, etc) is totally absent from today's elearning landscape. In the future, when massive amounts of learning resources are made available, the search engines should find the appropriate resources to achieve a given objective and attach the appropriate assessments to them. In order for this relation to be detected, it needs to be explicitly included in the document capturing the assessment.

## V. RECOMMENDATIONS TO ENHANCE INTEROPERABILITY

The case study described previously showed several venues that can be explored to increase or enhance the level of interoperability of QTI and assessment material in general. Although we are certain that perhaps all of them were discussed by the groups of experts that designed QTI, we think there is some value in reviewing them, specially after the significant changes that re-shaped the e-learning landscape in the last years.

### *A. Simplification of the specification*

The specification tries to cover with its structure a large number of scenarios with almost no adjustments on the scenario side. Although desirable, this is unrealistic. A simple specification capable of capturing all scenarios, even when some adjustments are needed should serve the purpose of interoperability better. Profiles are an example of this pragmatic approach. If the specification comes closer to these profiles (yet remaining unique), the need for a special solution for a specific context might decrease.

### *B. Focus on the essential aspects*

The aim of the initial specification was to cover perhaps too many aspects of the assessment scenario. One example of the vast set of possible situations is grading. The number of

possible grading policies is enormous. Providing a specification to describe all of them is equivalent to design a new programming language, or some new calculus notation. In our experiences, the basic functionality expected from the teaching staff when it comes to exchange assessment material is to be able to obtain a set of questions and answers from a given topic from a colleague. The instructions about how this material is graded is typically considered as highly dependent on the context of the assessment and therefore not perceived as essential.

Something analogous happens with the structure used to group a set of questions. In a certain course in an institution, a set of questions should be organized with a fairly special structure. If the course is divided into three topics and they need to be passed separately, the exam should reflect this situation. But when these questions are re-used in a different institution, these details are very unlikely to be re-used. Thus, a format leaving out this information would have a higher probability of being used.

A similar situation appears when assessments are transformed in "sequences". That is, rather than having a set of questions that are shown to the students all at once, there is some agent in charge of showing some of these questions, and even deciding which ones to show. This type of functionality, as in the case of the exam structure is highly dependent on the context. A better approach could be to adopt a "separation of concerns" approach. The set of questions are perfectly captured by a specification, and the different sequences to consider are the role of another specification that can be used in combination with the previous one. With this approach, institutions only interested in the plain exchange of questions would still be tempted to adopt the appropriate format.

### C. *Follow an incremental approach to deployment*

This aspect is related to the previous one. For the sake of completeness, it might be necessary to tackle some additional aspects surrounding assessments. In the previous section we argued that a clearly separated set of formalisms could be more appropriate. Together with this approach, a gradual deployment of these different specifications could greatly increase the level of adoption.

More precisely, if institutions have a simple path to question interoperability and make that an essential part of their day to day operations, they could be less reluctant to adopt a specification than enhances their procedure.

For example, if institutions exchange questions and answers with a simple specification, they would be more likely to adopt a specification to relate those questions with precise learning objectives that are already included in their courses. Providing a global and exhaustive solution in one single specification can be perceived as a much steep adoption curve and requires a more intense effort.

### D. *Acknowledge context without increasing the complexity*

Considering assessment as a stand-alone e/learning process is no longer acceptable. In the past, many courses would perform assessment procedures in platforms or contexts completely disconnected with the rest of a learning experience. But a proper design methodology needs a strong binding between learning objectives, activities, and assessment. Any specification to formalize assessment must be aware of this connections, although at the same time should minimize its impact in the complexity. Connecting an assessment with a set of objectives not only makes sense when designing a learning experience, but also allows that assessment to be easily found when searching for material in a much larger repository.

But the inclusion of these connections cannot mean a significant increase in complexity. Perhaps the most appropriate approach is to decouple these connections into its own formalism that may flourish on its own depending on the adoption, but at the same time maintains question management at a reasonable level of complexity.

## VI. CONCLUSIONS

A review of the most important issues in the context of assessment formats has been presented. Although the IMS QTI specification was conceived to offer a formalism to facilitate the exchange and reuse of assessment material among different platforms, the landscape after several years of evolution is somewhat confusing. QTI Version 1.2.1 has a significant level of support, although its intrinsic degree of flexibility turns into difficulties when it comes to implement a fully compliant import/export agent in a tool.

The scope of the specification as well as the number of structures and annotations that are considered optional offers a wide spectrum of possible solutions to encode assessment material. This richness has turned against interoperability as the complexity derived from this large number of solutions needs to be absorbed by the import/export agents.

A case study has been presented in which four institutions related to education (two universities and two e-learning commpanies) tried to exchange assessment material based on the IMS QTI specification. Three learning management systems were studied from the point of view of compliance with the process.

The study clearly shows the complex landscape derived from the flexibility of the QTI specification. Only two of the three platforms considered had the functionality to import/export assessment material in a common version of QTI. And still, the process could not be accomplished automatically because the resulting QTI files needed to be manipulated to accommodate certain specific features of the import/export modules of the other LMSs.

The recommendations to increase the adoption of QTI derived from this study are presented along four main strands|: simplify the specification, focus on the essential aspects of the assessment procedures, follow an incremental approach for deployment and acknowledge the context in which assessment is taking place without increasing the complexity.

## ACKNOWLEDGMENTS

This work has been partially funded by the iCoper Project (eContentPlus Best Practice Network No. ECP-2007-EDU-

417007), Learn3 Project ("Plan Nacional de I+D+I" TIN2008-05163/TSI), Flexo (AVANZA), España Virtual (CDTI, Ingenio 2010, CENIT), SOLITE (CYTED 508AC0341), and "Integración vertical de servicios telemáticos de apoyo al aprendizaje en entornos residenciales" (Programa de creación y consolidación de grupos de investigación de la Universidad Carlos III de Madrid).

## REFERENCES

[1] IMS Global Learning Consortium. Http://www.ims.org [accessed November 2009]

[2] IMS Global Consortium: IMS Question & Test Interoperability Specification. http://www.ims.org/question/ [accessed November 2009]

[3] Question Mark Corporation Ltd. Http://www.questionmark.com [accessed November 2009]

[4] ICOPER deliverable D6.1 Analysis of existing specifications and standards for assessment and evaluation and their usage in Europe. October 2009.

[5] IMS Global Consortium Common Cartridge. http://www.imsglobal.org/commoncartridge.html [accessed November 2009]

[6] Moodle. Open-source community based tools for learning. Http://www.moodle.org [accessed November 2009]

[7] Respondus Inc. http;//www.respondus.com [accessed November 2009]

[8] Hot Potatoes Version 6. Http://hotpot.uvic.ca [accessed November 2009]

[9] OpenMark http://www.open.ac.uk/openmarkexamples [accessed November 2009]

[10] SuML: Implementation of a multi-domain survey markup language. Barclay, M., Huq, S., Karras, B., & Lober, W. *Proceedings of the association for survey computing*, September 2002.