Assessing Assessment Formats: The Current Picture

Special Session: Assessing Assessment Formats

Israel Gutiérrez, Carlos Delgado Kloos, Raquel M. Crespo

Dep. Telematic Engineering Universidad Carlos III de Madrid 28911 Leganés (Madrid/Spain)

Abstract— **Student assessment plays a fundamental part in every e-learning process, where it can serve to check whether the learner has achieved the intended learning outcomes (summative assessment), but also as a means to aid in the learning itself (formative assessment). Nevertheless, there exist no formal standards to cover this type of content, just some specifications, such as IMS QTI. In this article, we present a study of formats for assessment and their usage in Europe. We also present a reference metamodel for assessment that covers the needs of all stakeholders in relation to this topic.**

Keywords - assessment formats, IMS QTI, reference metamodel, ICOPER.

I. INTRODUCTION

ICOPER [1] is a Best Practice Network that started in September 2008, funded by eContentPlus programme of the European Comission. As part of the ICOPER objectives, a reference model (ICOPER Reference Model) will be provided and some mechanisms to ensure involvement, cooperation and adoption of standards in the European educational framework. To accomplish this goal, the project will systematically analyse the specifications and standards available and in use, to draw conclusions on their validity.

In the context of the ICOPER project, an effort is under way to analyse assessment standards and specifications. This work has been carried out by the work package "Assessment and evaluation testbed", led by Carlos III University of Madrid. The analysis has focused on IMS Question and Test Interoperability (QTI) [2] because it is considered as *de facto* standard. Besides, this work package has other responsibilities like proposing a set of best practices in the scope of learning assessment, detecting and solving QTI interoperability problems by helping to complete the available tools to guarantee a robust exchange of assessment material [3].

This paper presents the results of the analysis of assessment formats and specifications, from the technical point of view, and their actual usage in Europe. As a result of this analysis, a reference metamodel for assessment content is proposed as a best practice in the assessment domain.

II. ASSESSMENT FORMATS AND SPECIFICATIONS

A. Methodology

The Model Driven Architecture (MDA) standard [4] has been used in order to analyse assessment specifications. Concretely, UML diagrams (what henceforth are called metamodels) have been used because they provide us with the following advantages as stated in [5]:

- Abstraction from secondary aspects such as syntax and concrete XML bindings.

- Better understanding of the format management.

- Development of systems easy to maintain and re-use (and therefore interoperable).

The criteria applied for the selection of the formats and specifications to be analysed are based on their relevance in the e-learning environment. Some additional formats used by ICOPER partners have also been included in the analysis due to their widespread deployment and experience as well as the possibility of an in-depth analysis of them granted by the consortium. The chosen specifications can be classified in two categories:

- **Assessment purpose formats**, whose main objective is the authoring of assessment resources. In this category, the studied formats are: IMS QTI (versions 1.2.1 and 2.1); formats belonging to Learning Management Systems (LMS), either open source systems like Moodle XML or commercial like Blackboard; and application specific like Hot Potatoes or OpenMark, an ad hoc format developed by the Open University of United Kingdom (OUUK).

- **General purpose formats**, whose original purpose is not learning assessment but can nevertheless be applied to it. The chosen formats belonging to this group are DocBook, used to write book and technical articles, FML, for describing Frequently Asked Questions, QAML, specification for question and answers, and SuML, used for writing surveys.

B. Comparative analysis of assessment formats

Once the assessment format metamodels were developed, a series of qualitative comparisons between them were carried out.

First, a set of key features was defined for assessment formats. The selection of features was based on the IMS QTI specification and the conceptual model that will be presented in Section IV:

- **Response and outcomes processing**: the possibility of processing the response given by the student in order to determine if it is correct or not; the processing of several question responses in order to get a global

result of the assessment. Response processing can be defined as an external process.

- **Metadata** capabilities: the possibility of storing the metadata of assessment items, sections and tests.

- **Hybrid question** management: the possibility of defining a hybrid question as a combination of a set of simple ones.

- **Correct response** indication: the possibility of indicating the correct response given a concrete question.

- **Multiple responses** related to one question: the possibility of defining more than one response to a given question (one correct and the others incorrect).

Table I summarizes the comparison of the analysed formats regarding the characterising features discussed above. The only common one is the use of metadata, but it is limited to a series of predefined fields like author or date in some formats. Most of them permit multiple responses to one question (only SuML and FML cannot represent it). The remaining features such as establishing the correct response, response processing or using a hybrid question are only supported by the assessment-purposed formats, that is, IMS QTI, Moodle, Hot Potatoes, OpenMark and Blackboard

TABLE I. KEY FEATURES IN ASSESSMENT FORMATS

Formats	Meta	Proc	M.R.	C.R.	Hybrid
IMS QTI	X	X	X	X	X
Hot Potatoes	X	X	X	X	X
MoodleXML	X	X	X	X	X
OpenMark	X	X	X	X	X
Blackboard	X	X	X	X	X
DocBook	X		X		
FML	X				
QAML	X		X		
SuML	X				

On the other hand, a series of question types have been selected in order to support the comparison between assessment formats. The selected question types have been classified according to the FREMA model [5] into the following sets:

- **Constrained response**: a question whose response is constrained to a space of solutions. That is, the student does not need to write any words, just to select an option from a set. Examples of this type of question are multiple choice questions (MCQ), multiple response questions (MRQ), true/false (TF) and matching.

- **Constructed response**: a question whose response is open, that is, it is constructed by the student. Examples of this type of questions are short-answer

questions, essays, fill in the blanks (FIB), crosswords, etc.

- **Miscellaneous/mixed**: questions that do not fit in the previous categories, like practice sessions, simulations, etc.

Taking into account this classification of question types, a set of them has been selected in order to make up a representative sample. The selected questions types, which have been collected from the assessment formats studied in the previous section, are as follows:

- Short answers: a textual answer consisting of a few words.

- Essay: a textual answer with unlimited or limited number of words that is not corrected automatically.

- Multiple choice question (MCQ): choose one option out of a list of possible answers. This type includes True/False (TF) questions.

- Multiple response question (MRQ): choose one, more or no option out of a list of possible answers.

- Fill in the blanks (FIB): complete missing words in a sentence or paragraph.

- Match: given two lists of terms, match each term on one list with one term on the other.

- Crossword: fill out a crossword using definitions of words in horizontal and vertical positions.

TABLE II. QUESTION TYPES IN ASSESSMENT FORMATS

Formats	Short	Essay	MCQ	MRQ	FIB	Match	Cross
IMS QTI	X	X	X	X	X	X	X
Hot Potatoes	X	X	X	X	X	X	X
MoodleXML	X	X	X	X	X	X	
OpenMark	X	X	X	X	X	X	
Blackboard	X	X	X	X	X	X	
DocBook	X	X					
FML	X	X					
QAML	X	X					
SuML	X	X					

Table II summarized the comparison of the formats regarding the types of questions supported. Short answer and essays are supported by all formats. Only formats developed for assessment purpose, however, allow multiple choice, multiple response, fill in the blanks or match questions. Crossword is a complex question type that Hot Potatoes supports and that can also be implemented in IMS QTI though in a more challenging way.

III. ASSESSMENT FORMATS USAGE IN EUROPE

As part of the analysis of assessment formats, a study of the usage of standards and specifications for assessments in Europe has been carried out. In order to accomplish this task, two different methodologies could have been followed:

- A quantitative analysis of a larger number of institutions using a statistical study

- An in depth analysis of a representative sample of institutions using both quantitative and qualitative methods.

The second methodology fits the objectives of the ICOPER project and can take advantage of the possibility of analysing, in depth, the institutions inside the consortium. Even more importantly, it is more appropriate for determining, analysing and understanding the underlying causes supporting the evidences found.

Figure 1 shows the distribution of use of the diferenent assessment formats, being the most spread ones IMS QTI, Blackboard and MoodleXML. The IMS QTI specification is considered as a *de facto* standard thus it is not unexpected to see its usage in several cases. Conversely the use of Blackboard can be justified because it is considered as one of the first e-learning tools. With respect to Moodle, (the most used e-learning tool currently) it also appears in the first position of this list. In the studied sample, it can be stated that the IMS QTI specification is used at least as frequently as proprietary formats that belong to two important tools in the e-learning scenario: Blackboard and Moodle. But considering all proprietary and ad-hoc formats as a single set, the ratio of usage of IMS QTI is rather low revealing a lack of acceptance of the specification. This statement is confirmed in [6].
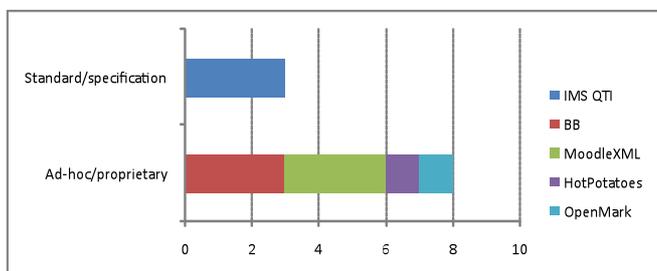

Figure 1. Usage of assessment formats

IV. REFERENCE METAMODEL FOR ASSESSMENT CONTENT

A. Proposal of a reference metamodel for learning assessment content

Due to the deficiencies detected during the metamodels analysis described in Section II and the analysis of required concepts in the assessment environment that will be presented in other paper of this special session, it has been considered necessary to propose a reference metamodel for assessment content.

Three parts have been identified in the assessment formats: presentation, processing and content. This model is focused on assessment content (it is intended to keep it simple to facilitate its adoption). It covers the assessment resources (content) and results (feedback and grade) from the assessment conceptual map. It is designed to provide assessment interoperability, so it avoids ambiguous definitions and duplication of information (such as some of the deficiencies found in QTI, as explained in the paper about interoperability in this special session).

The proposed reference metamodel accomplishes all key features discussed previously except processing, because this service can be externalised. In the same way, the proposed metamodel covers all questions types considered in the qualitative assessment formats comparison. Thus, the reference metamodel can be regarded as a minimum set of entities that accomplish the key features of existing assessment formats. This set is minimum in order to foster simplicity and ease of implementation, which are interesting features from the point of view of interoperability.

The reference metamodel for assessment content comprises all features required for completeness and interoperability in the learning assessment process. The basis of this metamodel is the IMS QTI metamodel and the generalized model created by Zuzana Bizonova in her PhD thesis [7] that is a study of a set of tools that have an assessment module, i.e., Claroline, Moodle and OLAT.

The initial version of the reference metamodel presented some deficiencies, the main one being the lack of support for learning outcomes different of knowledge, i.e., skills and competences. As competence-based learning is one of the ICOPER main objectives, it is necessary that the proposed metamodel cover these types of learning outcomes too. A new version of the model has thus been developed to fix such deficiencies.

The main concept of the reference metamodel, shown in Figure 2, is the assessment item. It represents an assessment entity that makes sense isolated. It is an abstract concept (like an abstract class in a object oriented programming language) that has three types of concrete realizations: question, task and assignment. Assessment items are composed of a definition, that is the wording of the question, and other metadata, like estimated duration of the assessment item or the author of this item.

Questions are used for assessing knowledge (like all the assessment formats studied previously) and they contain some entities related with this task: set of possible responses, correct response declaration and grading/feedback assigned to every possible response (in the case of constrained response questions).

Tasks target the assessment of skills, besides the knowledge that is implicitly used in such skills. They may make use of a series of hardware and/or software tools in order to perform the assessment. For example, ad-hoc software would be required for a simulation of the specific subject being assessed. Thus, tasks cover FREMA question type miscellaneous/mixed that were not covered by questions, although some constructed

questions can be also considered as tasks, e.g., essay assesses writing skills.
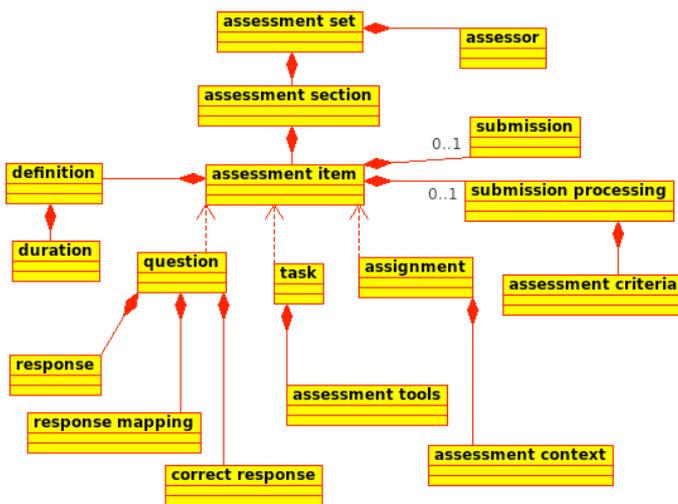

Figure 2.   Reference metamodel for assessment content

Finally, assignments are defined to evaluate competences (besides skills and knowledge) and require an assessment context. For example, work in groups can be assessed using a group of students that participate in the assessment process; this group of students is the context when a particular student is assessed.

On the other hand, student response (submission) is associated to assessment item, but this response can be or not be processed. It is also possible that this processing was defined as an external process. This processing is executed following a certain assessment criteria.

Finally, assessment items can be grouped in assessment sections, and sections in assessment sets that would correspond to complete exams. Given a complete example, it could be useful to define the evaluator in charge of this process, that is who assesses the student. The information of the assessor of an assessment set can be used to define the assessment method of this set, what is an important characteristic of this resource. For example, it can be useful as metadata of assessment resources in a repository in order to search automatic assessments (whose assessor is a computer).

B. Validation of a reference metamodel for learning assessment content

In order to validate the proposed reference model, model comparisons between the reference metamodel and the metamodels studied in Section II have been performed. A series of model transformations have been defined ad-hoc for this purpose. Thus, an algorithm has also been defined, which consists of an ordered set of transformation steps. After the executions of these steps, one model should be transformed into the other one; if so, it is demonstrated that they accomplish the same features.

The model transformations that have been defined to compare metamodels are as follows:

- **α-conversion**: an entity or attribute changes its name
- **attribute2entity**: an attribute changes its role to play like an entity
- **extension**: the addition of an entity or an attribute
- **relocation**: the change an entity or attribute position in the model

Once defined the basic transformations, an algorithm is executed consisting on a set of ordered steps. These steps are basic transformations, from a chunk of one model that accomplish a key functionality to a equivalent section in the metamodel that is being compared to.

This methodology has been used to compare the first version of the reference metamodel for assessment content to IMS QTI (version 1.2.1 and 2.1). The comparison results is that the proposed metamodel accomplishes the key features defined before, that QTI was supposed to accomplish based on its analysis. Thus, this reference metamodel was a simplification of QTI model maintaining the defined key features. It is assumed that these conclusions can be extended to the second version of the reference model because it has been built following the same premises of the first one.

V.   CONCLUSIONS

In this paper, a study of the most relevant assessment formats and specifications in present day e-learning systems is presented. A qualitative comparison has been performed among the studied assessment formats, following a list of defined key features and questions types. As a result of this analysis, a series of deficiencies have been detected in the studied formats, like the range of supported learning outcomes (limited to knowledge) and the lack of content interoperability. A study of specifications usage has been also presented, concluding that IMS QTI is not used as frequently as expected due to the detected problems [3].

Besides, a study of the concepts of learning assessment has been performed and, as a result of it, a concept map of this domain has been developed. This tool will allow us to know exactly the concepts used in learning assessment and the relationships between them.

Finally a reference metamodel for assessment content has been proposed, which solves the problems found in the specifications study. This metamodel is based on the comparison of assessment formats and the concept map developed previously. It should be also mentioned that the metamodel was defined having in mind assessment content interoperability.

As future work, it is planned the refining of the reference metamodel for assessment content in order to cover all necessary aspects in assessment domain. On the other hand, a data model is also planned to be created, based on the concept map and the reference metamodel, which will allow to carry out an actual implementation of this proposal.

A proof of concept will be executed in the Open ICOPER Content Space (OICS), a federated repository of assessment content, consisting on integrating and managing assessment

April 14-16, 2010, Madrid, SPAIN
IEEE EDUCON Education Engineering 2010 – The Future of Global Learning Engineering Education

1236

material. This implementation will be used as validation of the reference model presented in this article.

ACKNOWLEDGMENT

REFERENCES

[1]   ICOPER project website. http://www.icoper.org

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

[3]   Gutiérrez Rojas I., Agea A., Crespo R.M., Pardo A., Leony D., Delgado Kloos C., "Assessment interoperability using QTI", ICL2009 Conference, 23-25 Sept. 2009, Villach, Austria.

[4]   MDA from OMG. http://www.omg.org/mda/

[5]   Radenkovic S., Krdzavak N., Devedzik V., "A QTI metamodel". Proceedings of the International Multiconference on Computer Science and Information Technology. 2007.

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

[7]   Bizoňová Z., "Model Driven E-learning Platform" PhD thesis, May 2008.