Chapter 4

QTI: A Failed E-Learning Standard?

Michael Piotrowski
ZHAW Zurich University of Applied Sciences, Switzerland

ABSTRACT

The creation of good tests is time-consuming and expensive. Tests should therefore be reusable to ensure sustainability and to preserve investments and intellectual assets. This requires a standard, platform-neutral, vendor-independent interchange file format for tests. IMS Question and Test Interoperability (QTI) aims to be this standard. Almost a decade after the publication of the first version of QTI, even the interchange of simple multiple-choice tests between different systems remains problematic. In this chapter, the author presents a critical analysis of QTI. His conclusion is that QTI has failed to provide interoperability of questions and tests due to serious problems in its design.

INTRODUCTION

Assessment has always played an important role in education. Most, if not all, types of formal education use some sort of assessment, typically including a final exam to earn a grade, a degree, a license, or some other form of qualification.

Today, assessment is no longer restricted to grading at the end of a course (summative assessment), but it has been recognized that assessment is also useful for continuous monitoring and guiding of the learning progress (formative assessment), without being necessarily used for grading purposes (Boud, 2000).

Formative assessment, including self-assessment, can play a vital role in motivating students since it provides them with a way to judge their own competency level and allows them to track their progress. It also enables students to identify areas where more work is required, and to thereby remain motivated to improve further. Of course, this requires that students receive feedback as quickly as possible (Gibbs and Simpson, 2004).

Formative assessment also provides timely feedback for instructors, both with respect to the effectiveness of the course and the performance...
of the students; it thus helps to identify points that might need clarification.

For both groups, instructors and students, frequent testing is preferable. Case and Swanson (2002) argue that infrequent testing makes each exam a “major event,” with students investing much effort into preparation—they may even stop attending class to prepare for the exam. They also note that, with infrequent tests, students may be unable to determine whether they are studying the right material and with sufficient depth. Case and Swanson therefore conclude:

Though it may be more time consuming for faculty, frequent testing reduces the importance of each individual exam and helps students to better gauge their progress. (Case and Swanson, 2002, p. 116)

Assessment is always a time-consuming activity for instructors, especially if large numbers of students are to be assessed or, if assessment is frequent. This has motivated the development of technical devices to support assessment, starting with relatively simple mechanical devices in the 1920s and evolving to today’s computer-aided assessment (CAA) or e-assessment.

E-assessment is one of the fundamental elements of e-learning: Piotrowski (2009, p. 41) defines six activities that characterize e-learning platforms: Creation, organization, delivery, communication, collaboration, and assessment. Furthermore, of these six activities, assessment is the only one that is specifically educational; the other five activities are generic and not specific to e-learning.

The most frequently used form of e-assessment are multiple-choice tests. Multiple-choice tests have a number of practical advantages; in particular, scoring can be automated. This makes them especially attractive in e-learning settings, as it allows to make assessment available “anyplace, anytime.”

Creating high-quality multiple-choice tests, however, is challenging, especially if they are to assess higher-order cognitive levels, such as application, analysis, synthesis, and evaluation in the traditional taxonomy of Bloom (1956). Or, as Astin (1991) puts it:

While multiple-choice tests are indeed inexpensive to score, they are extremely expensive to construct: item writing is a highly refined and time-consuming art, especially if one expects to develop good items that are relatively unambiguous. (Astin, 1991, p. 148)

Ensuring the reusability, longevity, and platform independence of tests can mitigate the high costs of creation and can help preserve investments and intellectual assets when hardware and software change, thus ensuring sustainability. This requires a standard, platform-neutral, vendor-independent interchange file format for multiple-choice tests.

A number of standards aiming to promote interoperability and sustainability of e-learning content and e-assessment have emerged over the last decade. IMS Question and Test Interoperability (IMS QTI) is the best-known standard for tests. Most e-learning platforms today claim to support “IMS QTI.” And this is where the problems start: Which version of IMS QTI is supported? Does “support” refer to authoring or delivery of tests, or to import and export? IMS QTI has become one of the buzzwords in the e-learning community.

In this chapter, we will question this authority. When trying to implement IMS QTI or when trying to use IMS QTI for authoring tests, the user will soon become aware of serious shortcomings, which are a result of fundamental design flaws. Effectively IMS QTI is not able to fulfill the promise of interoperability. We therefore argue that IMS QTI is a failed e-learning standard. Unfortunately, this means that there currently exists no standard for interchanging tests. As a contribution to the development of a practicable future standard, we outline requirements an interchange format has to meet to actually play the role IMS QTI claims to play.