

# ***Modeling the Computing Based Testing domain extending IMS QTI: framework, models and exemplary implementations***

**Patricia Santos<sup>a</sup>, Davinia Hernández-Leo<sup>a</sup>, Mar Pérez-Sanagustín<sup>b</sup> and Josep Blat<sup>a</sup>**

<sup>a</sup>ICT Department, Universitat Pompeu Fabra, C/Roc Boronat 138, 08018 Barcelona, Spain // Tel.0034 935421428  
{patricia.santos, davinia.hernandez, josep.blat}@upf.edu

<sup>b</sup>Telematics Department, Universidad Carlos III de Madrid, Avda. Universidad 30, 28911 Leganés, Madrid, Spain  
mmpsana@it.uc3m.es

## **Abstract**

Using technology to create new types of questions is important, but also exploring new types of test representations or analyzing the most adequate technological device to support a specific assessment activity based on tests. This paper presents a conceptual model that defines and relates three dimensions to be considered for advancing on Computing-based Testing (CBT): the Question-item, the Test and the Activity. The framework assists in the categorization and design of advanced CBT scenarios and opens a new domain for questions and tests supported by ICT. Besides, the paper proposes two models for the technological design and implementation of the Test and Question-item dimensions. The models represent an extension of the IMS Question & Test Interoperability standard. They are platform-independent models (PIM) that formulate the elements that can be mapped to platform-specific models (PSM) for implementing advanced CBT solutions. To show the relevance, value and usability of the contributions, the paper also describes the application of the framework and PIMs to design and implement three exemplary CBT scenarios with their corresponding CBT-PSMs. Within the global scope of the CBT conceptual model, the first scenario shows how question-items can be enriched with specific multimedia information (web maps). The second scenario illustrates another particular case of the CBT domain where a test can be augmented with real physical elements. And finally, the third scenario describes an implementation that combines advanced question-items and tests using virtual worlds.

## **Keywords**

Computing-based Testing; Authentic learning; Formative assessment; Modelling Framework; Question and Test Interoperability

## **1. Introduction**

The use of Information and Communication Technologies (ICT) enables system developers and practitioners to rethink learning, teaching and assessment strategies. Particularly the use of automatic tests (or objective tests) in assessment provides advantages such as: improving the interactivity within the learning contents, presenting a question, obtaining a response, evaluating a response, providing a mark and answering with feedback automatically (Bull et al., 2002). However, despite the benefits of using automatic tests, the way of using traditional types of questions (e.g. Multiple Choice, Multiple Response or Fill in the blank) difficults the assessment of higher order skills such as: problem solving, problem exploration, collaboration, creativity, discovering rules, developing effective strategies, spatial or time perception, among others. Generally, these types of questions lack of complexity (Mayotte, 2010) and do not adequately capture the goals of an educational curriculum, fostering teachers' frustration (Bennett, 1998; Ridgway & McCusker, 2003). In this paper, when we refer to item's complexity we adopt the definition proposed by Luecht and Clauser (1998): *"item complexity is the result from features of stimulus, the nature of the responses elicited, and from the range of permitted examinee actions or activities"*.

Bull and McKenna said in their Blueprint for assessment (Bull & McKenna, 2004) that: *"CAA (Computer Assisted Assessment) is a common term for the use of computers in the assessment of student learning"*. But currently, some researchers state that the use of handheld computers will transform the learning and assessment practices (Dearnley et al., 2009; Thompson, 2005). This is the reason why this paper uses the term "Computing" instead of "Computer" when referring to Computing-based Testing (CBT) scenarios. Devices such as smartphones, PDAs, consoles or supporting technologies such as NFC or RFID tags, GPS and Bluetooth or Web 2.0 applications should be

considered and used in CBT activities (Santos et al., 2011). Technology Enhanced Learning (TEL) researchers are studying the benefits of using advanced technological solutions in education. Some of them are also analyzing the use of diverse devices (not only computers) to support learning in different spaces (beyond the classroom) in order to create enriched learning experiences (Hwang & Chang, 2011; Pérez-Sanagustín, Hernández-Leo, Nieves, & Blat, 2010; Oblinger, 2006). Recently, some assessment researchers explore the possibilities of using mobile phones for testing students (de-Marcos et al., 2010; Triantafyllou, Georgiadou, & Economides, 2008; Zhang, Wills, Gilbert, Warburton, & Bacigalupo, 2010). These solutions benefit from assessing students everywhere. However these solutions do not consider the physical space as a learning environment that can affect the way of answering the test and that can produce new assessment scenarios. We claim that taking into consideration the intrinsic characteristics of a device (for instance a mobile phone) opens up new possibilities for designing different types of assessment activities based on tests. These characteristics influence how students interact with a test and on the users' information that will be collected. Using the appropriate technology increases the possibilities of creating new types of questions, tests and assessment activities based on tests more adapted to real life tasks. As a consequence, the frustration that many teachers have when they design a test can be avoided because they have the possibility of creating more authentic assessment.

The use of ICT enables the design of new assessment solutions, allowing the creation of more complex scenarios based on tests. Creating advanced assessment scenarios should involve the use of technology to perform activities which would be impossible or very expensive to reproduce using traditional methods of assessment. ICTs offer the possibility of using simulations, managing a big quantity of updated and richer information, increasing the interaction with the information and making the student more participative in the assessment process (Conole & Warburton, 2005). Bennett (1999) indicated that the addition of multimedia resources in the creation of questions helps the teacher to assess more tasks. Recently, Boyle and Hutchison (2009) stated that the more technological resources you use, the higher skills and more sophisticated tasks can be evaluated. The JISC report "Effective assessment in digital age" (JISC, 2010) claimed that technology has to be used to create authentic assessment. This means that students should be able to demonstrate and practice their skills like in real life tasks. Then, as these authors claim, the use of technological resources allows the creation of more complex question-items. As a consequence, the assessment activities that are created can be more aligned with the current learning objectives of the educational curriculum.

The proposals of these authors have one common issue, they propose using computers in assessment based on tests for creating new interactions for question-items. We agree that using computers to represent the content and the type of a question-item is essential for advancing in this field, but it is also important to consider the characteristics of the appearance of the test itself. And also, it is necessary to take into account the educational elements and the technology which can be used to foster the development of student's skills during the assessment activity. We describe two of the most representative current test formats for evidencing the lack of consideration of the test presentation. The first example is the latest version of the IMS Question & Test Interoperability (QTI) v2.1. QTI is considered de-facto standard for representing and constructing question-items and tests (Harchay, Cheniti, & Braham, 2010). This standard introduces the test as the element that has to be fulfilled with a set of characteristics in order to organize the information related with question-items (IMS, 2006). The second example is the model used by the course management system Moodle (Dougiamas & Taylor, 2003), which defines its own XML format for tests (Moodle, 2011) so that it is possible to import and export questions. Both formats consider the test as a container of questions and as a mechanism for having an interrelation of questions to give a final report result. However, elements such as the appearance and interaction which are important aspects considered in the creation of questions, are rarely considered when designing a test. We address this matter by proposing elements concerning the design of a test that can be useful for helping students to better comprehend and contextualize the collection of questions or for increasing the complexity / richness of a test.

In order to know the elements that assessment system developers have to consider when designing and implementing advanced tests it is necessary to propose a re-conceptualization of the CBT conceptual domain. The current QTI Information Model (IMS QTI, 2006) is limited and does not consider elements to represent advanced CBT solutions (Joosten et al., 2007). As Joosten also claims, assessment system developers need to work on the same assessment conceptual domain to understand what elements can be re-used or ex-changed. This assessment domain will enable the analysis, the design and implementation of advanced CBT scenarios. This paper proposes a conceptual model including the Question-Item, the Test and the Activity as the three dimensions to be considered for conceiving advanced CBT assessment. The framework provides criteria for understanding the relations among the three dimensions, and can be used as an instrument for designing advanced CBT scenarios. The Test and the Question-item dimensions include a set of elements that have to be modeled in order to implement systems. In this way, the paper also proposes two platform-independent models (PIMs) in the sense that the specific implementation of the test

and questions interaction contexts is not considered in those models. The models serve as the base to develop platform-specific models (PSMs) and their corresponding systems. In order to validate the framework and the models, three advanced CBT scenarios are analyzed with the objective of evidencing the importance of the different elements considered. The models are based on the current QTI Information Model and extended with new elements in order to support the implementation of advanced CBT scenarios. [The contribution of the paper is a CBT conceptual model to facilitate the design, modeling and implementation of advanced CBT scenarios and systems. Since assessment is often embedded in learning processes for formative purposes, the elements of the conceptual model have been designed considering the potential of current ICT with the aim of facilitating richer assessments, more authentic and aligned with the expected learning outcomes. As a consequence, the learning value of assessment is improved.](#)

The remainder of the paper is organized as follows: in Section 2, we present a literature review and a discussion of the state of the art in CBT. Existing taxonomies, frameworks and models are analysed and discussed in order to detect limitations and propose solutions for modeling the CBT domain. Section 3 and 4 consider the limitations and problem domain analyzed in Section 2 and proposes the contributions for modeling the CBT domain. Section 3 presents the CBT Framework, proposed to facilitate the design of advanced CBT scenarios; and Section 4 introduces two models, the CBT-PIMs. These models contain a set of elements which determine how the test and the question-item have to be computed. In order to show the value and usability of the contributions, the application of the framework and PIMS associated to three real CBT scenarios is detailed in Section 5. Finally, Section 6 includes the conclusions and the future work derived from the contributions of this paper.

## **2. Previous taxonomies, frameworks and models for CBT: the need of an extended approach**

Several ways of capturing the elements related with assessment based on test have been proposed. In this section we review and discuss the need of extending other authors' proposals in relation with the three dimensions tackled in this paper: the Question-Item, the Test and the Assessment Activity.

### *2.1 The Question-Item dimension*

Koch (1993) organized **the Question-Item dimension** considering their grade of innovation: 1) Traditional items, 2) Items with graphics, 3) Multidimensional tasks and 4) Situated tasks (tasks with real-world congruence). The difference between these categories is their grade of technological complexity. Parshall, Davey, and Pashley (2000) proposed five dimensions to represent innovative questions-items: 1) Item format, 2) Response action, 3) Media inclusion, 4) Level of interactivity and 5) Scoring algorithm. The dimensions can be considered as a set or separately. As Bennet identified (1998), in order to propose advanced question-items an author has to make a "re-conceptualization", this means that they have to think about: how to use technology to represent the question, how users can interact with questions and how the answers can be scored. Zenisky and Sireci (2002) explained that one of the most critical elements that assessment system developers have to think about is the format of the response of a question. These researchers stated that the more complex the way of answering a question is, the more complex the representation of the question is.

*From these proposals we can extract that technology affects in different aspects to question-items making them more complex. First, a question is more complex if it uses more technological resources for its visual representation. And second, as more complex is the interaction that a user has to perform for answering a question-item, more technological resources are used to represent and compute it.*

We use an example to clarify this statement. Imagine a question about Geography which asks about finding the city of Paris in a map. This question can be represented (1) using a static image of the European continent with a text asking about finding Paris. Or (2) using an interactive web-map (e.g. see Google, 2010) with the zoom, drag and satellite appearance functionalities activated, and the same text asking about Paris. In the first example, the student only has the possibility of doing a point over a position. The tasks that the student can do with the image are to observe and draw over it. In the second case, the student has the possibility of doing zoom in order to give a more concrete position. Also, students can drag the map to contextualize it in the correct zone. In addition, the satellite appearance offers the possibility of giving more details about the location, in this case Paris, adding new layers of 3D information. As we can see in the second case, where the representation of the question is more complex, also the way of answering the question increases its complexity.

## 2.2 The Test dimension

The visual and interaction aspects of the **Test dimension** are hardly considered in the literature. To understand better this problem we use three examples of well-known authoring test tools selected by the ICOPER project (2011): 1) Hot Potatoes (HotPotatoes, 2011), 2) Moodle quizzes (Moodle, 2010) and 3) the ReCourse editor (which integrates a section for editing QTI tests) (ReCourse, 2009). These tools enable managing the parameters of a test to organize questions. However, none of these tools include options for using media and changing the visual aspects of a test. The test that results from these editors is a group of questions organized in a table with a simple layout where questions are organized one below the other. The ReCourse editor is compliant with the de-facto standard for assessment, QTI. Concretely, ReCourse uses the QTI version 2.1. This version defines an Information Model (IMS QTI, 2006) which contains the classes and elements necessary for processing and rendering “items” (question-items) and tests. Part of the QTI Information Model is the content model that contains an element called “itemBody”. The itemBody contains the information (text, graphics, media objects, and interactions) that describes the content of a question-item and its structure. The body of the question-item is represented using style sheet information, or using the default style rules of the specific authoring tool. In the QTI v2.1 Information Model the test object is composed by the following elements: 1) The *navigation and submission mode*. The first one defines the paths that users can follow in order to answer the questions. The second determines when the response of a question is processed. 2) The *test structure* defines the organization of testParts, sections and questions. 3) The *time limits* establish the duration that users have for answering a question or a section of a test. 4) The *outcome processing* is the element that is in charge of computing the results obtained in a question or group of questions. Finally, 5) the *test feedback* is the group of outcome variables that can be shown to the user in order to know their final report result. As we can see, none of the elements consider the visual and interaction aspects of a Test.

*We propose that an Interaction Context element could be included to increase the complexity of tests, its visual representation and the way students can interact with the information, as well as providing a more concrete context where question-items can be located.*

The following two examples can help to better understand the importance of the *InteractionContext* element. 1) In a Biology subject a teacher is preparing a test that contains questions which ask about human anatomy. The questions have relation with specific parts of the human body. Then, instead of presenting the test (group of questions) in a blank style sheet, technology can be used to show the questions over a 3D graphical representation of the human body (e.g. see Google, 2011). In this case, the context (the 3D human body) can help students to better reflect about the questions and at the same time to learn anatomy meanwhile they respond and explore. The teacher can locate the questions over the different organs of the graphical body. 2) A teacher of History wants to prepare a test about the Hellenistic period. This period had important events across different years. The teacher is interested in asking about some specific events, but also wants the students to understand the evolution of the historical facts during this period. This can be solved using technology to represent the test with an interactive time-line (see an example of time-line in SIMILE, 2003) where the events in the time-line are questions. The time-line can be useful for the teacher to organize the questions, and for the students to understand the evolution of the Hellenistic period. These examples aim to show that in some situations *changing the representation of a Test (and how to locate questions in a context) and increasing the interaction among students and the test, can be useful to provide information that students can use to answer and to understand better the relations between the questions and the test.*

## 2.3 The Assessment Activity dimension

Finally, in order to decide the better way of representing questions and tests, it is important to have in mind the **Assessment Activity** dimension. Bloom and Krathwohl (1956) defined a taxonomy for modeling the cognitive domain. This taxonomy was reviewed by Anderson and Krathwohl (2001) and is known as the revised Bloom’s Taxonomy. The taxonomy organizes *skills* into six different levels. The Bloom’s Taxonomy can be used to know the type of skills that teachers can assess regarding a specific learning objective. Level 1: Knowledge; Level 2: Comprehension; Level 3: Application; Level 4: Analysis; Level 5: Synthesis; and Level 6: Evaluation. The higher a level is, the most sophisticated is the task that the user has to do in order to answer the question. This means that the participation and implication of the student interacting with the information contained in a test or/and in question-items would be higher. As Bull & McKenna (2004) and McBeath (1992) stated it is assumed that a test can be only used to assess the three or four first levels of the Bloom’s taxonomy. However these researchers indicated that depending on the design of the questions and the test, the six levels can be assessed. Biggs (1999) stated that the educational curriculum, the teaching, the learning and the assessment activities have to be aligned. In the same

direction, Joosten et al. (2007) proposed a model for new and traditional assessment types. Although the model goes beyond assessment based on test, some of the issues that they identified are useful for discussion. For instance, they claimed that assessment has to be integrated with learning and instruction because assessment addresses complex traits of students. Joosten explained that when an assessment activity is designed it is necessary to think about the traits that have to be assessed, this information is saved in the *Assessment Plan*. This plan is used to construct the *Assessment Scenario* (which determines the mandatory and optional *Units of assessment* for a user, their sequence and the time schedule). In Joosten's model a Unit of assessment is formed by *Items* interpreted in a broader sense (an Item can be a question, an exam, a delivery of an essay...).

*We agree that to correctly design an assessment activity, the learning aspects have to be considered. However, we state that it is also important: 1) to reflect about which the most adequate technology as well as 2) to identify the learning environment where the test has to be answered.*

The ideas presented by the ICOPER Project can be used to answer the first point. The aim of the project is to identify best practices for the adoption of specifications and standards across Europe. One of their works is devoted to course evaluation. The project proposes a concept map for TEL, which is part of the ICOPER Reference Model (IRM) (ICOPER, 2011). This conceptual map indicates that an assessment activity has to use assessment methods, resources and tools. By *assessment tools* they referred to authoring tools or engines for assessment. But we propose that this element has to integrate also *technological devices* and *ICT supporting tools*. Depending on the device selected it is possible to design one assessment scenario or another. In other words, it is not the same to use a mobile phone with GPS for answering geolocated questions during a route, than answering a test with a computer in classroom. This issue is addressed by the studies made by the FREMA (2007) reference model for the assessment domain and Wills et al. (2007). FREMA indicates that the delivery context is an important aspect that has to be included in our assessment scenario design. They distinguish between a *digital environment* (e.g., an assessment management system) or a *physical environment* (e.g., a classroom, a museum, among others). We propose that the characteristics of a *learning environment* have to be considered when designing advanced CBT scenarios. For example, when answering a test in a museum students need to collect information from the pictures. Depending on the learning environment the students can put into practice one type of skills or other. The FutureLab report of mobile technologies and learning claimed: "*Learning will move more and more outside of the classroom and into the learner's environments, both real and virtual, thus becoming more situated, personal, collaborative and lifelong. The challenge will be to discover how to use mobile technologies to transform learning into a seamless part of daily life to the point where it is not recognised as learning at all* (Naismith, Lonsdale, Vavoula, & Sharples, 2004)." According to this idea, we identify three possible types of learning environments: 1) *Distance* when the assessment activity does not depend of any real space and the users can interact with the test wherever they wanted. 2) *Face to Face*, when the students have to be located in the same space. And, 3) *In situ* when the questions which conforms the test of the assessment activity are related with a location, and students need to have context awareness. This last category can be divided in two sub-categories: 3.a) *in virtual situ* or 3.b) *in real situ*. As Herbert (2010) and Liu and Hwang (2010) stated, embedded technologies and augmented realities enable to explore and use physical and virtual spaces such as learning environments where the user has to have context-awareness to do correctly the activity. *In this paper we propose that it is necessary to distinguish between virtual or real spaces, when designing an assessment activity. This distinction helps to select the adequate technology and take profit of its characteristics to assess the student's skills correctly.*

This analytical review let us to propose a comprehensive approach that integrates the Question-item, Test and Assessment Activity dimensions, so that new extended test-based assessment approaches can be devised. The approach proposed includes a conceptual framework and two models for advanced CBT. On the one hand, the framework is proposed to understand the CBT conceptual domain, and the relations between the elements of this domain. The elements contained in the framework can be used to categorize existing CBT scenarios and also to devise new ones. On the other hand, two models for the Test and Question-Item dimensions are proposed to provide the information model elements that can be considered in the implementation of these dimensions. These models are PIMs that extend QTI with new elements independent of a specific implementation case. Assessment experts can use the CBT-PIMs to create the PSMs for a specific application case and implement the particular advanced assessment tools. The next sections present in detail the framework and the associated models.

### **3. Conceptual framework: CBT dimensions**

The conceptual CBT framework (see Figure 1) is graphically represented as a set of structures, elements and its relationships around three dimensions: the Assessment Activity, the Test and the Question-Item. An advanced CBT

scenario is designed with the aim of collecting evidences from the students for testing the outcomes about what they have learnt and providing the adequate feedback about their progress (ICOPER, 2011). The aim of the CBT framework is to facilitate a comprehensible representation of the elements and the relationships which have to be considered for designing advanced CBT scenarios.

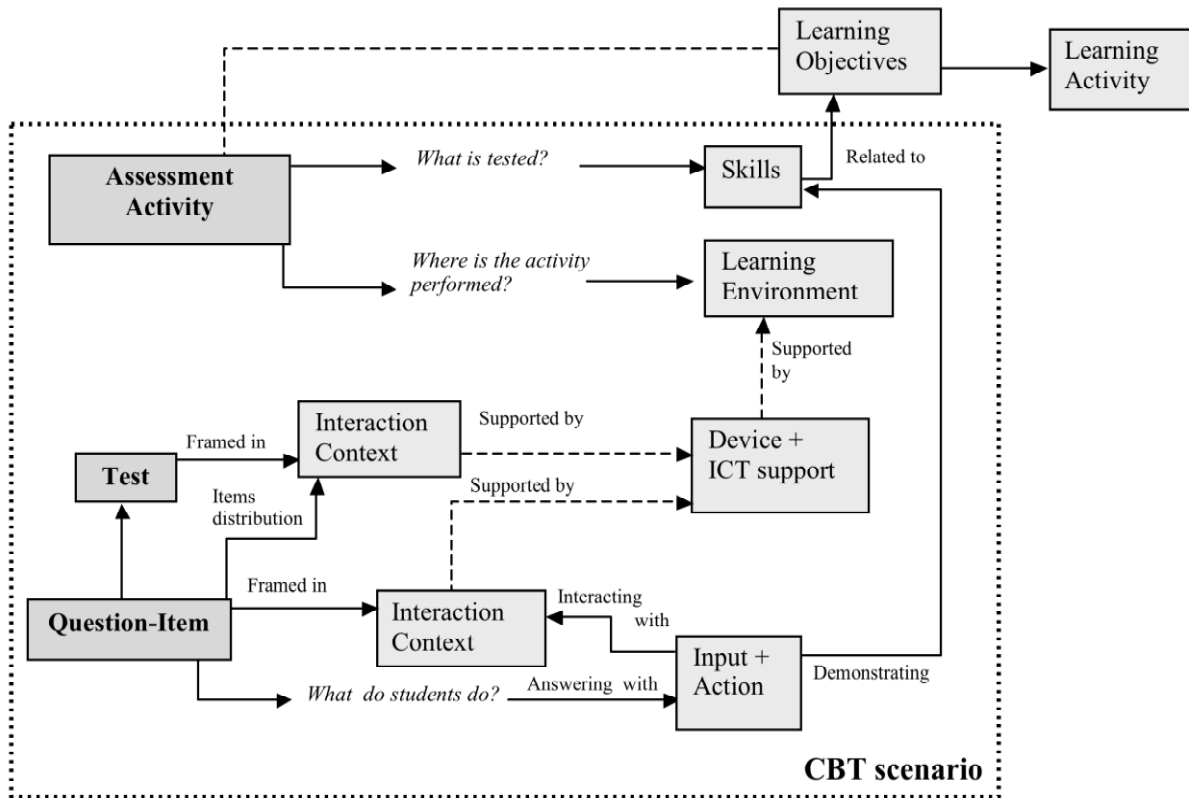


Figure 1. CBT Framework

First of all, when an advanced CBT scenario has to be designed it is necessary to think about: *what do we want to test?* The framework defines an **Assessment Activity** as the dimension that connects a *Test* and the corresponding group of *Questions-items*. By answering questions students have to put into practice and demonstrate their *Skills*. These skills have to be related to the *learning objectives* of a concrete *learning activity*.

As we have discussed in the previous section, it is important to select an adequate *learning environment* to perform an assessment activity. Herbert (2010) proposed that before selecting a learning space (what we called *learning environment*) it is necessary to identify the learning content and the assessment strategy that is going to be covered. Then, the best environment to support the educational issues can be identified. In the framework, the *learning environment* is proposed as the situation where the student has to be located to perform the assessment activity. Depending on the type of learning environment selected (*face to face*, *at distance* or *in situ*) the *technology* selected to support the *Test* and the *Question-item* will change.

We have discussed that advanced CBT solutions have to pay more attention in the characteristics of the **Test** dimension. The CBT framework proposes that an *Interaction Context* has to be considered to facilitate the understanding of a test. On the one hand, the interaction context has to facilitate the set of specific virtual or real objects to represent the topic assessed. As mentioned in the examples of the discussion the interaction context can be: a 3D representation of a body, a map, a room of a museum, etc. Students will have to use this information to answer correctly the questions. On the other hand this context has to be used to distribute and organize question-items. This organization can help students to understand the relation between the different questions and the whole test.

The **Question-item** is the third dimension of the framework. This dimension can be also represented with their specific *Interaction Context*. In this case, students have to interact with the context conformed by a set of objects

(virtual or real) to answer the question. To illustrate this, and following with the same examples used in the Test dimension, the set of interactive objects used to represent a question-item could be: a 3D representation of the heart, the view of the street-map of Barcelona town-planning, or the Sunflowers' painting by Van Gogh. The information selected to represent the interaction context of a question has to facilitate the assessment of specific students' skills. This means that it has to be considered how the students will interact with the information to answer the question. It is very important to design well this interaction because it provides information about the skills developed by the students to provide their answer. Then, it is necessary to detect the *input* (product) introduced by the student as the answer. But also, the *actions* (process) followed by students to provide the answer. We claim that the input and the action facilitate capturing more concrete information of the answering process. This information will be used to provide more adequate feedback.

Both, the interaction context of the test and the question-item have to be supported by *Technological devices* and *ICT supporting tools*. The characteristics of the devices and/or ICT supporting tools have to: 1) be aligned with the characteristics of the learning environment where the activity will be performed, 2) facilitate the representation and use of the objects of the interaction context of the test and the question-items and 3) enable putting into practice and assess the students' skills. As several authors discuss, different technologies that can be used to support assessment depending on the characteristics of the CBT scenario. For instance, Ridgway & McCusker (2003) discussed the use of computers in assessment to assess new educational goals. Elliott (2008) identified the benefits of using web 2.0 services in assessment. Dearnley et al. (2009) explained the importance of using mobiles for testing students and Motiwalla (2007) proposed a framework with the factors that have to be considered to create mobile learning environments. The technology selected has to provide support for the students in order to facilitate putting into practice and to comprehend the tasks they have to perform. In addition, the technological solution has to facilitate teachers the assessment of the students' skills in a specific learning environment.

#### **4. Modeling the Test and Question-Item dimensions of the CBT Framework**

To approach the implementation of the Test and Question-Item dimensions we adopt the main concepts of the Model Driven Architecture, which propose to formulate platform-independent models (PIMs) illustrating a set of elements for representing a system without any platform dependency. PIMs have to be translated to platform-specific models (PSMs) using concrete technologies and programming languages systems (Kleppe, Warmer, & Bast, 2003), so that they can be implemented in actual systems. A PSM is the model of a system that combines the PIM specifications and the characteristics of the specific platform selected.

In CBT the elements of the Test and the Question-item dimensions have to be implemented in order to use them in a technology-supported Assessment Activity. This section proposes two PIMs, one for the Test and another for the Question-Item dimension. The models contain elements which belong to the Information Model of QTI, and new ones based on the CBT framework. The QTI model was selected to represent some elements of the models because it is considered the de-facto standard for assessment (ICOPER, 2011). The other elements are proposed as an extension to QTI in order to support advanced CBT scenarios. The CBT-PIMs are focused on the characteristics that CBT systems have to operate, while the technical specific aspects derived from the platform will be captured on the corresponding CBT-PSMs.

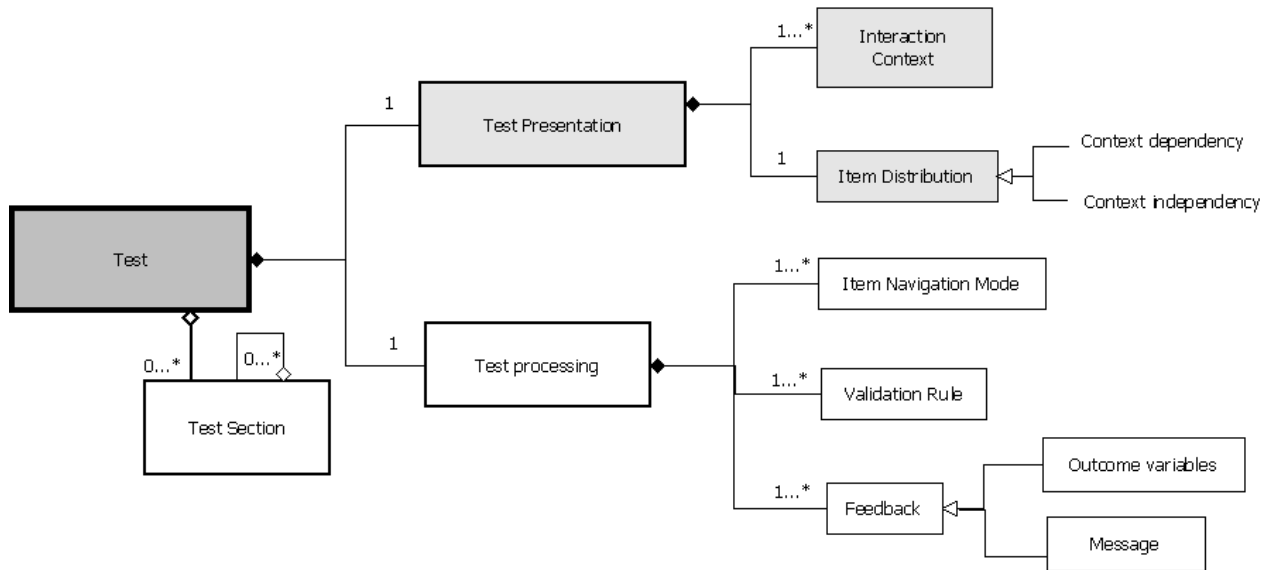
The CBT-PIMs of this paper provide the elements to guide assessment system designers and developers in the creation of their own CBT-PSM. From the CBT-PIMs and considering the details of the specific platform selected these users can create the CBT-PSMs and implement the specific system.

To facilitate discussion and comprehension, the models do not provide details about the corresponding XML implementation. The extension of QTI is based on the elements contained in its UML Information Model. For this reason the CBT-PIM elements and relations are represented using the Unified Modeling Language (UML).

##### *4.1 The CBT-PIM of the Test dimension*

As the CBT framework includes, the information of a Test is based on the information related to the *Assessment Activity* dimension. The following model contains the necessary elements for processing and representing advanced Tests. A *Test* is the computational representation of a group of *Question-items*. But the test is also the dimension in charge of giving a common context to understand the relation of the set of question-items. Also, the test is the dimension which has to facilitate the comprehension of the learning goal that is going to be assessed.

This model extends QTI but is independent of the *Interaction Context* implementation. It is a PIM from the point of view of the *Test Presentation*, meaning that it can be particularized into a PSM when a concrete interaction context is implemented in a Test.



**Figure 2.** Test dimension elements of the CBT-PIM. The new extended elements are represented in grey, the elements that follow the QTI standard in white.

The **Test Presentation** depends on an *Interaction Context* which determines the *Items Distribution*. As the CBT framework captures, the interaction context is the set of virtual or physical objects that has to facilitate the understanding of a *Learning Objective* and has to promote putting into practice the student's *skills*. In the *interaction context* the question-items have to be organized, following (or not) a criteria. When the characteristics of the context are essential for answering correctly the questions of the test, a *context dependency* is established. For instance, in the case of *in situ* activities the *Learning Environment* (see the Framework, Figure 1) influences the interaction context. On the contrary, if the context does not have any influence over the questions, the test has a *context independency* relation. The next examples can help to understand better the Test Presentation elements: (1) the *interaction context* of a test about Geography can be a web map where the questions are located (*ItemsDistribution*) in different coordinates of the map. The questions asking about Africa are located over this country and the same with the other countries. Students can explore the map to understand better the questions and their organization (*context dependency*). (2) The *context* of a test in Mathematics can be a blank html table where the questions are located in different rows, in this case the context is not important and students do not need to visualize certain visual characteristics to answer the questions (*context independency*). And (3) in an Art subject the physical space of a museum can be the *interaction context* of the test. Different QR-code tags can be put under the pictures with a link to a question-item. In this case the objects of the physical environment help the students to discover the questions and to understand their relation with other pictures of the museum (*context dependency*). Also the structure of the environment itself (a museum with rooms and pictures) is used to locate the questions.

The elements of the **Test Processing** are defined according to part of the QTI information model (IMS QTI, 2006). This set of elements determines how the test has to be computed. It is composed by: (1) the *Item Navigation Mode*, which establishes the predetermined order that the user has to follow in order to answer the question-items. For example the navigation mode is *Adaptive* when an item appears whether another concrete item has been answered previously. Or *Sequence* when the items have to be answered following a specific order. The *Validation Rule* is the algorithm or set of actions that compute the total of actions done by the student in the Test. As Macdonald (2003) indicated, two aspects can be assessed: (1) the process, which is the number and types of actions conducted by the student for answering the test (these actions have to be the sum of the interactions performed in the interaction context of the test and in each question-item); and (2) the final product which is the computation of the total of answers. As a result of applying the validation rule the final outcome variables are created (e.g. score, number of correct answers, number of incorrect answers, and level of skills...). Finally, the *Feedback* is the information element that determines which *outcome variables* and *message* are showed to the user. The outcome variables are

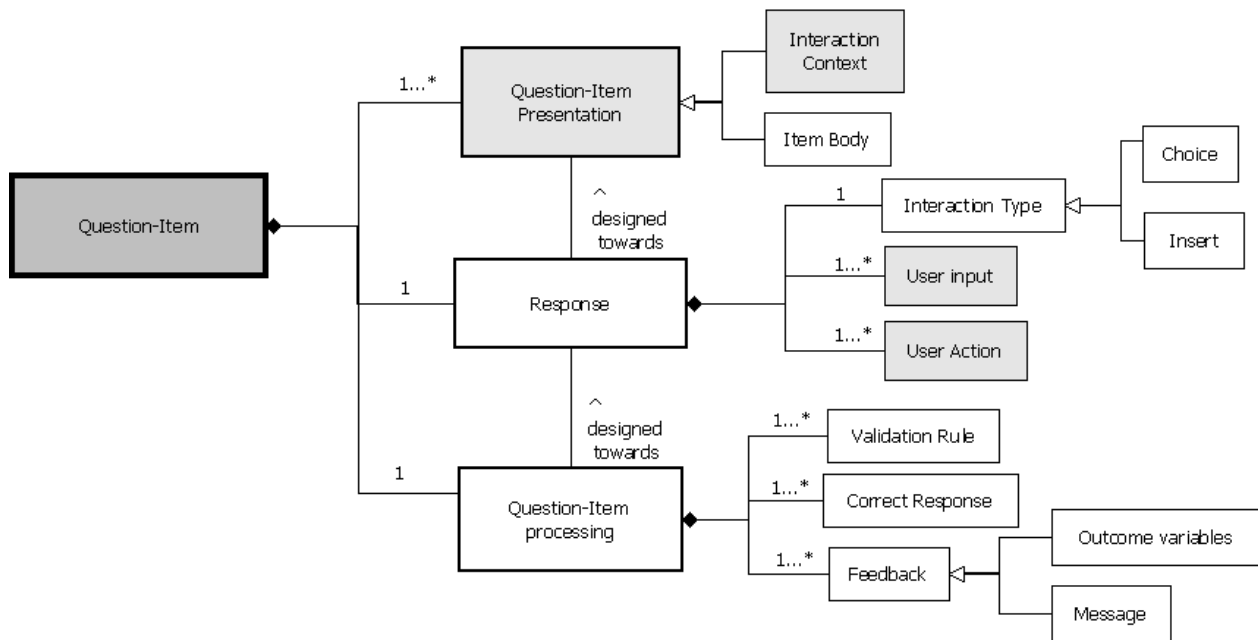


used to create the test result report. The message can be textual or multimedia information associated with the answers.

Finally, a test can be divided into different *Test Sections* in order to organize similar question-items.

#### 4.2 The CBT-PIM of the Question-Item dimension

A *Question-item* is the computational representation of a question. As Parshall et al. (2000) stated, the use of technology enables creating advanced question-items, which are impossible to reproduce using traditional methods. The Question-Item PIM also follows the IMS QTI but, again, it is independent of the interaction context. In this case the *Interaction Context* has an influence on the other elements: the *Response* and *Question-item processing*. The Platform Specific Model (PSM) will be based on a concrete interaction context. The selection of a specific interaction context will affect the implementation of the corresponding *Response* and *Validation Rules* of the question-items.



**Figure 3. Question-Item dimension elements of the CBT-PIM. The new extended elements are represented in grey, the elements that follow the QTI standard in white.**

The Question-Item has to be designed considering the sophisticated task and group of skills expected to be assessed. For instance: draw frontiers between countries, put anatomic organs in their correct position or solve a mathematic equation with two unknown factors.

The **Presentation** of a question-item is an element that affects directly the complexity and its comprehension. The Presentation element is composed by: (1) The *Interaction Context*, which has the same characteristics that the *interaction context* of a Test applied to a question. In this case the interaction context has to be useful to put into practice the necessary skills for answering correctly a question. (2) The *Body*, which contains the text of the question and its possible responses.

The element **Response** is composed by: the *Interaction type* that defines the category of the question-item. Based on the organization of types of question-items proposed by Bull & McKenna (2004), two big categories of interactions are proposed: a) *Choice*, these are closed questions where the correct answer/s is/are contained in the *context* of the question and the user has to find it doing some action. For example: Multiple Choice, Yes/no, Multiple Response, Order, Match, etc. And the category b) *Insert*, open or a closed questions where the user has to introduce the answer because it is not contained in the *context* of the item. For example: Fill in the blank, Open question, Calculate, Draw, etc. The QTI information model includes an *Interaction type* element which can be organized as we have proposed.

The CBT Framework, based on Macdonald (2003) work, proposes that besides the *Interaction type* it is necessary to distinguish between the final *product* and the *process* followed by the student to answer a question. The *User Input* is the product that the user has to introduce in order to answer the question. This is the element which is computed as the answer of the question. In order to submit the user input the user has to perform a number of actions. For instance *User Actions* are: drag pieces of information of the question, draw a line, select between different choices, click on a position, among others. The *user action* determines the grade of participation of the user in the process of answering a question-item. The task of answering (which is the sum of the *UserInput* and the *UserAction*) allows assessing the student's *skills*. When the user action requires a higher participation of the student, the assessment activity, the test and the question are more sophisticated (Boyle & Hutchison, 2009) and, as a consequence, the skill level is higher. The Response element also includes the *Answer* which is the list of possible responses and their values.

The next examples can help to understand better the question-item presentation and response elements: 1) the *interaction context* of a question-item about Geography can be an interactive map of the Iberian Peninsula where the text of the question (*ItemBody*) asks "Draw the frontier between Spain and Portugal". Students can explore the map to understand better the question, they have to draw over the map (*UserAction*) and provide a line (*UserInput*) as the answer. 2) In a Botany subject a real plant of a natural park can be the *interaction context* of a question. In this case the question can be a traditional text multiple choice question (*ItemBody*). However, to answer the question correctly it is necessary that the students interact with the real plant doing a photograph or measuring the length of its leafs (*userActions*). This action will allow the students to select the correct choice (*userInput*). Both the *Presentation* and the *Response* elements (and its corresponding sub-elements) depend on the *InteractionContext* selected as the most adequate to represent the CBT scenario. The selection will determine the creation of a platform-specific model (PSM).

Finally, the elements of the **Question-Item processing** are also based on part of the QTI model. The *Validation Rule* is the algorithm that checks whether the *user input* is the same than the pre-established *correct Answer*. Once the item is computed a *Feedback* is sent to the user. The feedback can contain outcome variables such as a message, the score, the correct response, etc.

## 5. Three Advanced CBT scenarios and implementations and the corresponding CBT-PSMs

In order to evaluate the modeling approaches proposed in this paper, this section includes an analysis of existing implementations of the CBT-PIMs into PSMs, the resulting software tools and their use in experiments. The technological setting used in these scenarios can be described with the CBT conceptual model. The analysis of the results obtained and the experience acquired during the realisation of the cases were used to refine the conceptual model. As a result, three CBT-PSMs were developed. Each PSM describes the technical aspects derived from a specific system to implement the visual representation and interaction of tests and question-items extending QTI.

These implementations were used and evaluated in real educational contexts with students and teachers to understand the educational benefits of the CBT scenarios implemented. The first scenario, QTIMaps, analyses the educational benefits of implementing advanced question-items following the CBT conceptual model proposed. The second scenario, QuesTInSitu, studies the elements related with the Test dimension and their benefits in assessment. Finally, the third scenario, Wonderland-QTI, shows the results obtained by a third party-institution after following the CBT Framework and PIMs. This case shows how advanced representations of questions and tests can be combined in the same CBT scenario. As shown in previous research (Navarrete et al., 2011; Santos et al., 2011 and Ibañez et al., 2011), these scenarios are relevant and represent advanced approaches to assessment.

This paper explains how the Framework has to be used to think about the elements and relationships that are necessary to have into account when the CBT non-trivial scenarios used in the experiments are designed. The paper also shows that within the global view provided by the CBT conceptual model, the PIMs proposed can be used: independently depending on the dimension of the model (Question-Item or Test) that is more relevant to a specific case (as shown with the first two systems/ experiments, QTIMaps and QuesTInSitu), or together when both dimensions are important and need to be combined (as shown with the third scenario, the Wonderland-QTI scenario). These implementations permit us to demonstrate new ways of designing CBT scenarios using advanced ICTs. Moreover, these experiments have been used to show the applicability of the CBT conceptual model in real assessment scenarios.

### 5.1 QTIMaps: extending QTI questions with web map content

This scenario presents how the elements of the CBT Framework and the CBT-PIM of the Question-Item can be followed to design and implement adequate advanced Question-Items to cover the specific educational goals of a curriculum. The scenario explains a real case of development of new types of questions for assessing geographical and spatial higher order skills.

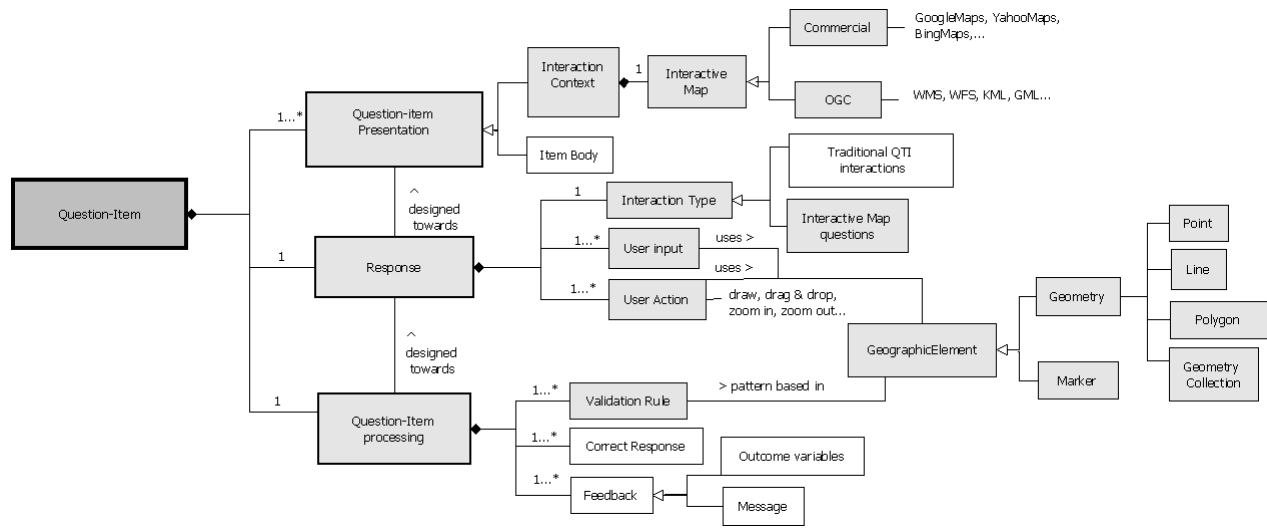
The CBT Framework was used, jointly with a teacher, to design the following assessment activity. The educational curriculum of the Spanish Geography and History subject (Spanish government, 2006) indicates that students have to put into practice the following skills: (1) to identify, localize and analyze, in different scales, the basic elements that characterize the physical environment; (2) to understand the territory; (3) to identify, localize and understand the basic characteristics of the geography diversity of the world and the biggest geoeconomical areas; (4) to identify and localize in the time and in the space the relevant processes and historic facts of the world history; and finally (5) to search, select, understand and relate verbatim, graphic, iconic, statistic and cartographic information, from different sources (books, media, and information technologies). These skills are very difficult to be assessed using traditional types of questions. However, current web map applications allow practicing these skills interacting with multimedia geographical information. Considering the three dimensions of the framework, the dimension with higher relevance in this case is the *Question-item* dimension. Then an interactive web map can be used to represent the *interaction context* of questions. This interaction context will enable the students to put into practice the skills previously cited. Students have to interact with the maps doing *actions* that are typically supported by web map applications. In addition, it is necessary that they introduce *inputs* over the map as final answer to the questions. In this case, considering that the interaction context of the question-items has all the information that students need to practice the geographical and spatial skills. Consequently, it is not necessary to use a complex *interaction context* for representing the *Test dimension*. Question-items can be distributed in the test following a format of html table, one question below the other.

This assessment activity can be performed in a *learning environment* at distance (e.g. at home) or face to face (e.g. at classroom). A good *device* to support the activity is a computer (PC or laptop). Also, it is necessary to use a web map application, *ICT support* (e.g. Google Maps, Yahoo! Maps, etc.), to represent the interaction context of the question-items.

Once the requirements of the scenario and the elements of the three dimensions are completed following the CBT Framework, the CBT-PIMs can be used to design specific solutions. In this case a particularization of the Question-item CBT-PIM was proposed as: QTIMaps PSM. The QTIMaps PSM is the result of applying the PIM having into consideration the characteristics of Google Maps (Google, 2010). The model is represented in Figure 4, the elements which follow the QTI standard are in white, and the new elements are highlighted in grey. Although, QTI v2.1 has the possibility of using graphics (e.g. maps) as *interaction context* of a question, the standard does not contemplate interacting in a more authentic way with maps. Google Maps was selected as a provider of maps because it has an open Application Programming Interface (API). In the QTIMaps solution the *Question-Item Presentation* consists of the following elements: (a) the *itemBody*, it is the textual information contained in a QTI question. And (b) the *interactionContext*, an *InteractiveMap*, it is a map from GoogleMaps (other web map tools could be used).

Then, it was necessary to extend QTI with new types of interactions. In this case these new interactions are called *Interactive Map questions*. Taking into account the skills that had to be tested, the students needed to put into practice *User Actions* such as: zoom, changing the appearance of a map, drawing lines to differentiate regions, etc. And introducing *User Inputs* such as: a click over a position, a marker in the correct order, etc. These user actions are typically supported by web map applications, including Google Maps.

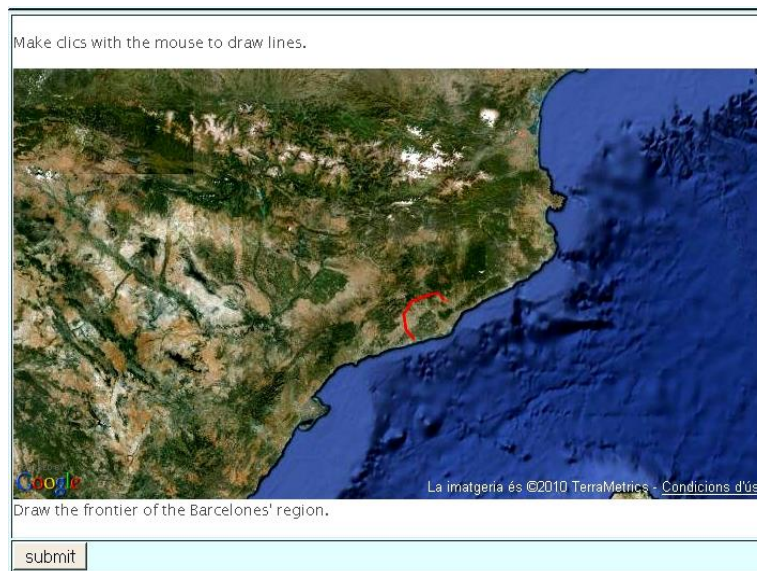
The users interact with the web maps doing actions and inputs using *GeographicElements* provided by the web map application. The answers are corrected automatically by the QTIMaps solution following specific algorithms and patterns (*Validation Rule*) especially designed to support the actions and inputs of the students with the *GeographicElements*.



**Figure 4. CBT-PSM QTIMaps.** *The new extended QTIMaps elements are represented in grey, the elements that follow the QTI standard in white.*

A partial reference implementation of the QTIMaps model, covering the whole lifecycle of edition and runtime was developed. The details and evaluation of the system are presented in Navarrete, Santos et al. (2011). The paper is focused on showing the results of applying the QTIMaps system in various experiments including an authentic assessment activity with 23 students and their teacher in a high school. The evaluation shows educational benefits going beyond assessment, such as enhancement of students’ motivation and memory reinforcement, related to the fact that students had to interact actively with web maps to answer the questions.

The QTIMaps model covers other additional assessment activities framed in different educational scenarios where georeferencing is relevant. Figure 5 shows an example of question designed by a teacher, considering the interactions of the QTIMaps solution. In this question, students have to visualize a map of the world in which they had to use the zoom tool, find the city of Barcelona and draw an approximate frontier of the region.



**Figure 5. Advanced Question-Items, example of QTIMaps question**

The CBT Framework and the CBT-PIM of the Question-Item dimension was used to design and implement the QTIMaps scenario, the PSM and the system presented in this section. The design of the QTIMaps PSM and the evaluation of the system in authentic assessment activities enabled us to test the educational benefits and usability of the CBT conceptual model proposed. The use of multimedia maps as interaction context of questions offers new and

more authentic ways to assess relevant geographical skills, such as spatial thinking. The QTIMaps model is applicable and supports a variety of assessment scenarios, and it can be used in different educational disciplines apart from Geography. The results of this case demonstrated that considering the presentation, and response elements of a Question-item, it is possible to design more authentic and enriched questions augmenting the participation of students in the process of answering questions. QTI can be extended with enriched interactive contexts combining them with traditional types of questions or creating new ones.

### 5.2 QuesTInSitu: Assessment in situ, contextual awareness for representing tests

In this case the Framework and the PIMs are used to analyze the importance of providing more complex representations of the Test dimension. Specifically, the section explains how the CBT Framework and the CBT-PIM of the Test dimension can be used to design and implement an assessment activity for assessing higher order skills *in situ*. This type of assessment activities cannot be practiced without mobile devices, and on the contrary to other m-test activities, they are dependent of specific locations. The words *in situ* mean that questions only can be answered when users are correctly placed in a specific location. In an assessment *in situ* activity students have to interact with the physical environment putting into practice transversal skills such as exploration, spatial and observation besides of the specific skills related with the content of the test.

First the CBT Framework was used with real teachers to design authentic assessment *in situ* activities. In this case, the *learning environment* (e.g. a route in a city, in a natural park, etc.) influences directly in the design of the *interaction context* of the *Test*. The questions have to be associated to real positions and distributed as a route in the physical environment selected (e.g. a route in a city). Considering these factors two types of *technological devices* have to be used to support the activity: computers to design the assessment activity, and mobile phones with GPS to answer the geolocated questions *in situ*. The framework is used to reflect and take advantage of technology to advance especially in the representation of the test, but the questions can be traditional types of question-items. The innovation in this type of activity falls on the *presentation* of the test (a route using a map and questions associated to real spaces) and the *user actions* that the students had to do to answer the questions (observe the environment, touch the material of a building, talk with the citizens, etc.). Then, the way of representing the test in this case increased the sophistication of the questions and enabled the assessment of higher order skills.

This scenario innovates especially in the Test dimension and shows how the CBT-PIM Test can be used to design a specific PSM for supporting assessment in situ: the QuesTInSitu PSM. The QuesTInSitu PSM (Figure 6) represents the distribution of geolocated QTI Questions-Items as routes (*Item Distribution*). The QTI questions are associated to real GPS coordinates. In this case two types of *interactionContexts* are used to represent the Test dimension: (1) On the one hand, an interactive map from Google Maps is used as *virtual* interaction context where questions are associated to GPS coordinates. (2) On the other hand, students have to be *physically* in a *real* space (a city, a natural park...) which corresponds to the real area of the map used to geolocated the questions.

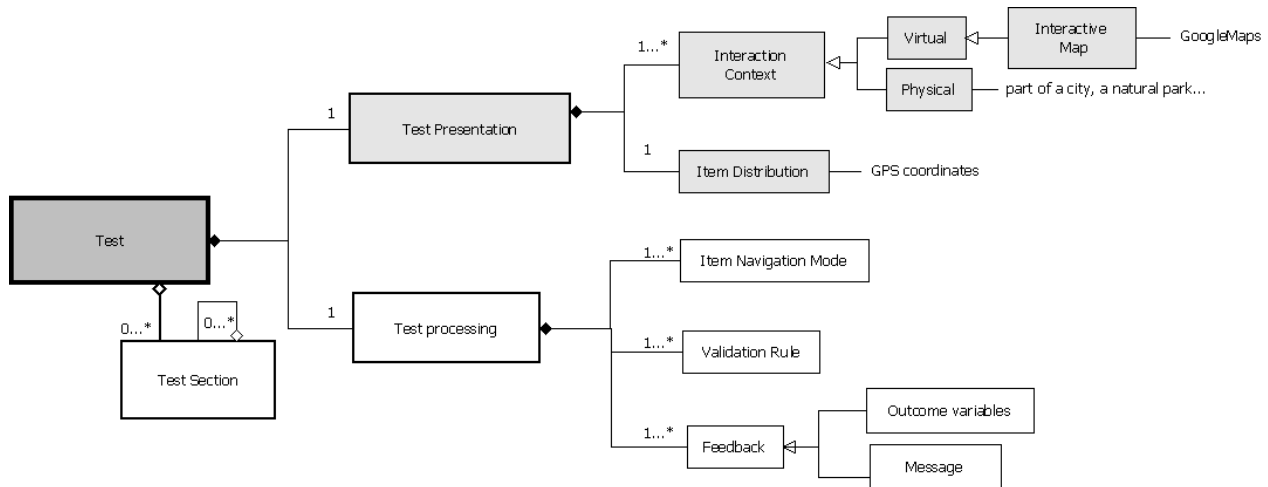
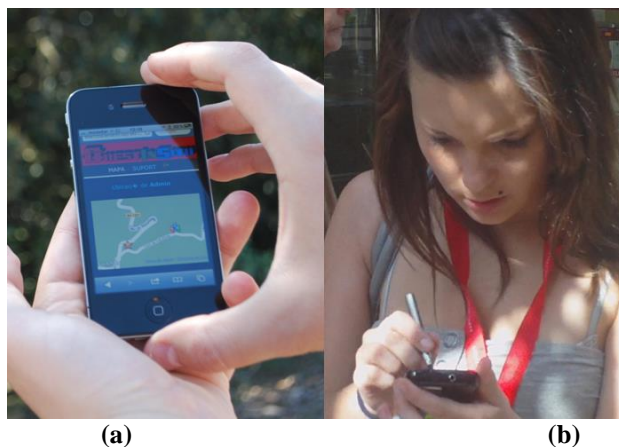


Figure 6. QuesTInSitu-PSM. The new extended QuesTInSitu elements are represented in grey, the elements that follows the QTI standard in white.

The QuesTInSitu PSM was followed for implementing a corresponding reference implementation demonstrating that the model is useful for implementing specific systems. The QuesTInSitu system is detailed in Santos et al. (2011) and evaluated in a real assessment in situ activity with students and teachers showing educational benefits.

The assessment activity was titled Discovering Barcelona, focused on assessing the town-planning skills of the students in situ. The teacher had previously designed six different routes (*tests*) with associated QTI questions (Multiple Choice, Multiple Response and Yes/No). The students followed their route using a smartphone (see Figure 7.a) and when entering in a question-area the text of the question appeared automatically in the mobile screen. The questions had to be answered in the established positions, using QuesTInSitu mobile (see Figure 7.b), because their content was related with the environment of the area. If the students were not located in the correct “situ” (place) they would not be able of answering correctly the answer.



**Figure 7. QuesTInSitu (a) Looking for questions (b) Student answering a question in situ**

The main results obtained in the experiment showed that QuesTInSitu gives a valuable support for teachers and students in the performance of assessment associated to open real world environments. Representing the Test as a route of geolocated questions, receiving the question and feedback in situ and using smartphones for doing the activity facilitates students to put into practice their explorative skills and improve their motivation.

This section shows that the characteristics of a learning environment (the city in this case) have to be considered when designing an assessment in situ activity. The learning environment had an effect on the presentation of the Test and the distribution of Question-items. In addition, due to the characteristics of the environment, specific technology (mobile phones with GPS, and web-maps) were used to support the representation of the Test dimension. The association of GPS coordinates to QTI questions enables the creation of an innovative CBT scenario. The elements of the QuesTInSitu PSM enable to extend QTI for implementing real assessment in situ activities with a test represented as a route of QTI geolocated questions.

### 5.3 Wonderland-QTI: virtual worlds for representing tests and questions

The last scenario shows how the CBT Framework and the PIMS can be used to include a virtual world as *interaction context* to create advanced tests and question-items to assess Literature and language skills. When students learn a new language, they have to read and write using this language correctly, but also they have to know how to respond to real circumstances (not only talking but also doing actions). For instance: when a person, who is learning English, goes to England s/he has to know how to buy tickets on the subway. A virtual world offers the possibility of representing any environment: a museum, a world described in a novel, a historical period, etc. In a 3D world avatars can perform human actions such as: moving objects, talking with other people, walk, etc. This opens up the possibility of using the space of the virtual world and the avatars for simulating *learning environments* where the avatars have to be conducted by students (Dillenbourg, Schneider, & Synteta, 2002). The avatars have to *interact with the context* (objects of the world) making *user actions* that simulate human real tasks to demonstrate their *skills*. The CBT framework is used to obtain the design of the scenario to assess literature and language skills. In this scenario, the *interaction context* of the *Test* is a set of virtual objects used to represent the world. In this world questions are located in specific positions. A set of objects around a specific position determines the *interaction*

context of a *Question-item*. Avatars (conducted by students) have to interact with these objects in order to answer a question. This type of assessment activity is especially designed to be performed in a *distance learning environment*.

This CBT scenario is an example that shows how a third party institution can use the proposed conceptual framework to use technology for representing specific assessment activities for assessing language skills. The CBT-PIMs were used in this stage, the third-party institution decided to use the virtual world Wonderland (Wonderland, 2010) for representing the interaction context of the Test and Question-Item dimension. This is modelled in the Wonderland-QTI PSM (see Figure 8). The Wonderland-QTI -PSM extends the QTI standard with new elements to make possible the use of virtual worlds to assess students' skills. The 3D space of the Wonderland virtual world is used as the *interaction context* where the *Question-items* of the *Test* are distributed and assigned to a specific coordinate of the world. Students can use the 3D graphical information of this space to understand the goal of the Test. Each question-item is represented with a specific set of *WonderlandElements* (3D objects and/or avatars) as *interaction context*. The *itemBody* of the question can be presented as text or as a message given by other avatar. The student has to solve the question doing some *actions* (moving the 3D objects, walking to a certain position, touching an avatar) and sending the *input* (the selection of a specific object, the click over a position, etc.) as the answer. Finally the question-items are processed applying specific validation rules that are based in patterns that consider the *WonderlandElements* that form the question. In this case, the *Feedback message* can be enriched with *WonderlandElements*. Instead of showing a textual message, an avatar or 3D objects can appear after answering a question as a message of response.

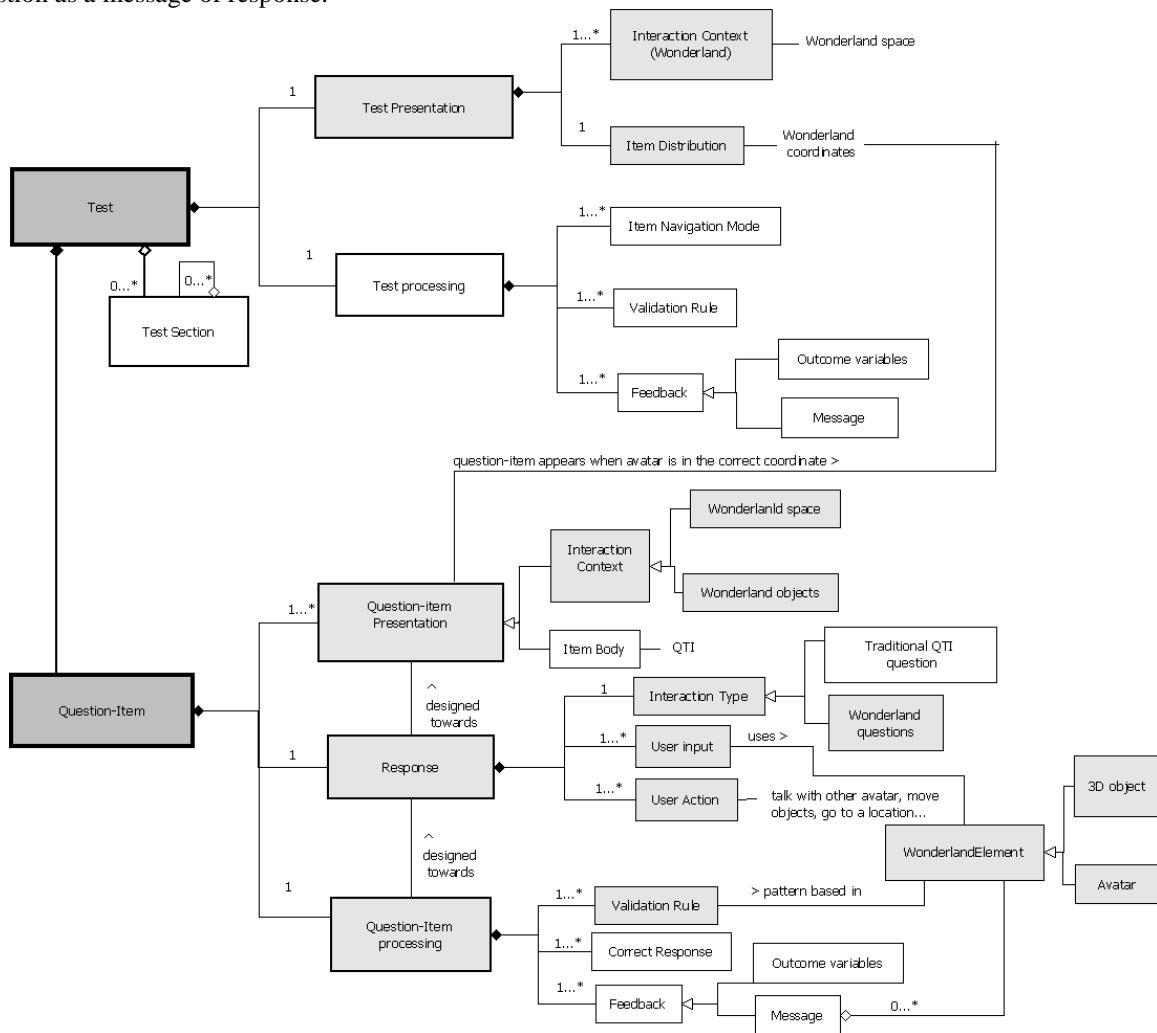


Figure 8. Wonderland-QTI PSM. The new extended Wonderland-QTI elements are represented in grey, the elements that follow the QTI standard in white.



The result of the Wonderland-QTI PSM was the system Wonderland-QTI, implemented by the third-party and detailed in Morillo, et al. (2010). These authors explain in Ibañez et al., (2011) how Wonderland-QTI was used for simulating the setting of a science fiction novel and implementing an assessment activity based on the extra-curriculum of the engineering studies. The activity was carried out by 12 students. The Test and Question-items were represented, using Wonderland, as an immersive and motivating *interaction context*. In this context, students had to interact with the different objects in order to demonstrate their knowledge about the novel in the way they would have to do so in the real world (see Figure 9). The *interaction context* imitates some scenarios and characters of the novel. The world makes use of *ICT support* such as natural text chatting with synthetic characters, textual tagging of virtual objects, automatic reading of texts, and the integration of a 3D mouse in learning sequences in order to exploit the capabilities of 3D virtual worlds.

The results evidenced that participants felt the sense of physical immersion, they were very motivated of answering a test interacting with a virtual world and this interaction helped them to understand better the novel.



**Figure 9. Avatar demonstrating his language skills**

The importance of this application case deals with how a third-party institution has followed the CBT Conceptual Model for implementing a system adapted to their particular circumstances. This case exemplifies how a rich interaction context (in this case the 3D virtual world) can be used to enhance traditional types of questions. The Wonderland-QTI enhances traditional QTI questions representing the choices with 3D objects or areas, and the user has to interact with these elements in order to answer the question. The question-items can be distributed in the world, and the avatars have to interact doing *user actions* with the objects in order to provide a *user input*. As demonstrated by other authors before (de-Freitas & Oliver, 2006, Jong, Lee, Lee, & Shang, 2010; Shang, Jong, Lee, & Lee, 2008), virtual worlds are a good option to simulate a learning environment where students can demonstrate their skills. Assessment system developers can take the Wonderland-QTI PSM as a reference model for representing assessment activities into virtual worlds.

#### 5.4 Comparison of the three CBT scenarios

As a summary of this section, Table 1 relates the main elements contained in the CBT Framework, with the main characteristics of the three CBT scenarios (QTIMaps, QuesTInSitu and Wonderland-QTI). The table also includes the main results of these scenarios. The aim of the table is to provide a comparison of the main characteristics of each scenario and their main results for an easier understanding of the differences and similarities among these CBT scenarios.



Table 1. Summary of the main elements of the CBT Dimensions of the three CBT scenarios

CBT Dimensions Scenarios	Assessment Activity	Test	Question-Item	Main results
QTI-Maps	<p><b>Skills:</b> to analyse in different scales, the elements of the map; understand the territory and relate verbatim, graphic, iconic, statistic and cartographic information; to recognize, to identify, to localize and to appoint the geography elements</p> <p><b>Learning environment:</b> at distance (e.g. at home) or face to face (e.g. at classroom)</p> <p><b>Device + ICT Support:</b> Computers, GoogleMaps</p>	<p><b>Interaction context:</b> (<i>non especial interactive context</i>) a simple html table, one question below the other</p>	<p><b>Interaction context:</b> interactive multimedia web-map (e.g. Google Maps)</p> <p><b>User-action:</b> students can select choices, can draw over the map, also they can do clicks over locations. Also the users can do zoom in and zoom out of the map, can drag elements, they can change the appearance of the map (satellite visualization, road map, a mixture of both)</p> <p><b>User-input:</b> a choice, a line and a click</p>	<p>The use of multimedia maps as interaction context of Questions-items offers new and more authentic ways to assess relevant geographical skills, such as spatial thinking, augmenting the participation of students (user-actions) in the process of answering questions</p>
QuesTInSitu	<p><b>Skills:</b> to interact with the physical environment putting into practice transversal skills such as exploration, spatial and observation besides of the specific skills related with the content of the test</p> <p><b>Learning environment:</b> open spaces with 3G and GPS coverage (e.g. a route in a city, in a natural park, etc.)</p> <p><b>Device + ICT Support:</b> Mobile devices with 3G and GPS, Computers, GoogleMaps</p>	<p><b>Interaction context:</b> test is represented as a route of geolocated questions using a multimedia-map and questions associated to real spaces</p>	<p><b>Interaction context:</b> (<i>non especial interactive context</i>) traditional QTI question-items</p> <p><b>User-action:</b> walk to find the questions, observe the environment, touch the material of a building, talk with the citizens, etc.</p> <p><b>User-input:</b> a choice</p>	<p>Representing the Test as a route of geolocated questions, receiving the question and feedback <i>in situ</i> and using smartphones for doing the activity facilitates students to interact with a physical environment (user-actions) in order to answer contextual-dependant questions and put into practice their explorative skills, improving their motivation</p>
Wonderland-QTI	<p><b>Skills:</b> learning a new language means to read and write this language correctly, but also to know how to respond to real circumstances (not only talking but also doing actions)</p> <p><b>Learning environment:</b> a virtual world</p> <p><b>Device + ICT Support:</b> Computers, Virtual world Wonderland</p>	<p><b>Interaction context:</b> the test is represented through a 3D virtual world simulating a specific environment, questions are located in different positions of this world</p>	<p><b>Interaction context:</b> 3D objects of the virtual world</p> <p><b>User-action:</b> simulation of human real tasks: observe and take objects, talk with avatars, walk to specific locations, etc.</p> <p><b>User-input:</b> a choice, a click</p>	<p>Virtual worlds can be used to represent tests and question-items with enriched interactive contexts (3D information) enabling the simulation of real-life tasks</p>

## 6. Conclusion

This paper proposes a conceptual model of three main dimensions to be considered for advancing on Computing-based Testing (CBT): the Question-item, the Test and the Activity. On the one hand, a CBT framework is proposed to assist practitioners and technicians in the categorization and design of advanced CBT scenarios, which can be supported by different types of devices and software platforms to enable the assessment of higher order skills. On the other hand, two platform-independent models (PIMs) provide the elements for computationally describing the Test and Question-item dimensions. The CBT-PIMs are extensions of the existing IMS QTI Information Model. Concretely, these extensions include a set of elements that enhance the *visual representation (presentation, interaction context and item distribution)* and *interaction (user input and user action)* of tests and question-items. The resulting CBT-PSMs combine the characteristics of a specific platform (used to represent the new elements extended in the models) with the traditional elements of QTI.

Apart from studying approaches available in the literature, we specifically analyze three existing advanced assessment scenarios (Navarrete et al., 2011; Santos et al., 2011 and Ibañez et al., 2011) conceived according to the CBT conceptual model. The application of the conceptual model to these scenarios and their analysis illustrate the suitability and usability of the Framework and PIMs to lead to specific PSMs and their corresponding systems and evaluated experiments.

The first scenario, QTIMaps, illustrates the implementation of new types of question-items. Students interact with the questions not only for answering them but also for understanding the information visualized in the question, putting into practice spatial and geographical skills. With QTIMaps we show how QTI can be extended with geographical interactions provided by Google Maps. This case used the CBT-PIM of the Question-Item dimension to define the interaction context, the user action and input using Google Maps. The result was the QTIMaps-PSM and the corresponding implementation. The solution combines the benefits of using QTI with the geographical interaction context for questions and functionalities that a web map provides.

The second scenario shows that it is possible to consider the Test dimension and their related elements for implementing advanced CBT scenarios. The results obtained in the experiments done with the QuesTInSitu system show the potential benefits of these elements. The CBT Conceptual Model and the participatory design with teachers enabled us to design the QuesTInSitu-PSM and its corresponding implementation. QuesTInSitu demonstrates that QTI tests can be represented as a route of QTI geolocated questions, offering a new way of representing tests and new types of CBT activities. This system enables the creation of traditional QTI questions that are associated to real coordinates where the physical environment is used as the interaction context. Students need to have physical context awareness in order to answer the question and demonstrate their exploratory skills.

In the third scenario, the CBT-PIMs were used to extend QTI with the Wonderland virtual world as interaction context for representing advanced tests and question-items. As a result the Wonderland-QTI PSM and system was implemented. This scenario shows that the CBT Conceptual Model is useful for developing advanced CBT scenarios and systems envisaged by other authors. A third-party institution developed a PSM and a system where a virtual world is used to implement advanced QTI questions and tests to assess literature skills.

The scenarios and systems implemented in these experiments represent exemplary contributions for the CBT domain and QTI. The modeling effort, through the proposed QTI extensions, shows that the scenarios present general traits, which can be generalized (and reused) in other CBT scenarios. We have not provided details about the XML implementation, to avoid getting into excessive technicality, and making more difficult to follow the discussion thread. However, as indicated in the exemplary scenarios, the mapping of the models into XML and its turning to experiments has been realised. In that sense, mapping the CBT models proposed to XML has been tested, and, even the exemplary scenarios could be turned into reference implementations if the extension proposals were included in QTI.

As concluding remarks, from the results of the experiments and the implementations, we can state that the experiments and the resulting implementations evidence that the CBT Conceptual Model proposed enables the design and implementation of advanced CBT scenarios and tools, extending the scope of the IMS Question and Test Interoperability. Moreover, we contend the models and the exemplary scenarios can support other authors to envisage new advanced CBT scenarios. The QTI standard was used for computationally representing part of our proposals and it was extended with other external services for implementing the additional elements proposed in the CBT-PIMs. After evaluating the resulting implementations in a diverse range of real assessment scenarios evidence the educational benefits and advantages of advanced CBT assessment. These CBT scenarios enable teachers to design tests and questions for assessing transversal (such as exploration, teamwork, ubiquitous skills) and specific higher order skills. In addition, the activities implemented enhance the students' motivation. The resulting CBT-

PSMs are exemplary models that show the combination possibilities of QTI with other specific technologies. Assessment designers and developers can use these models to implement their own advanced solutions. The CBT-PSMs presented are reference models than can be used in the future to include the extensions proposed in the QTI Model. In conclusion, we can state that the implementation of the scenarios and the results obtained after the evaluation evidence the benefits of using the CBT conceptual model for creating advanced CBT activities. However, it is necessary to evaluate the use of the model with a significant number of technicians and practitioners in order to detect the facilities and limitations that they have reading the CBT Framework and the CBT-PIMs. This evaluation will be use to validate the elements and relations of the CBT conceptual model.

We propose as future work to include the extensions proposed by the CBT models in a future version of the QTI standard. This extension could be use to implement a wide range of advanced CBT activities. The implementation of more examples of scenarios will show the potential and flexibility of the proposed elements. The experiments presented in this paper used technologies such as computers, smartphones, GPS and multimedia graphics to computationally represent the questions and tests. However, other technologies also could be used. Future implementations will let show that question-items and tests can be supported by different technological devices opening up the possibility of supporting new types of assessment activities based on tests.

In addition it is necessary to explore automatic transformation of the CBT-PIMs to CBT-PSMs. In this proposal we have adopted the concepts of Model Driven Architecture of PIM and PSM. However the potential of this architecture for the automatic transformation of models has not been explored and represents an interesting future work. In future research, an authoring tool with suggestions for automatic mapping from CBT-PIM elements to specific CBT-PSM elements (according with previous selected requirements) could be implemented.

Finally, we plan to work in the development of the corresponding CBT tools (e.g., authoring tools, engines) to support advanced CBT scenarios. In the cases of QTIMaps or Wonderland-QTI specific software has been implemented for managing the extended elements, not included in QTI, and representing questions with enriched interactive data. The scenario of QuesTInSitu provides an authoring tool which supports the design, creation and implementation of routes (tests) of geolocated QTI questions. Assessment developers have to take into account that the success of their implementations in real assessment scenarios will be possible if the systems can be used by their final users. More research and evaluation regarding the implementation of user-friendly authoring tools for creating CBT activities should be done.

## Acknowledgements

This work has been partially sponsored by the Spanish Ministry of Science and Innovation Learn3 project, TIN2008-05163/TSI and the Spanish Ministry of Science and Innovation EEE Project TIN2011-28308-C03-03. The authors would also like to thank the rest of the members of the GTI research group and the GAST research team.

## References

- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching and assessing: A revision of bloom's. taxonomy of educational objectives: Complete edition, New York : Longman.
- Bennett, R. E. (1998). Reinventing assessment: Speculations on the future of large-scale educational testing. A policy information perspective. Princeton, NJ: Policy Information Center, Educational Testing Service.
- Bennett, R. E. (1999). Using multimedia in large-scale computer-based testing program. *Computers in Human Behavior*, 15(3-4), 283-294.
- Biggs. (1999). Teaching for quality learning at university. Buckingham: Society for Research in Higher Education and Open University Press.
- Bloom, B., & Krathwohl, D. (1956). Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university examiners. Handbook 1: Cognitive domain. New York: Longman.
- Boyle, A., & Hutchison, D. (2009). Sophisticated tasks in e-assessment: What are they and what are their benefits? *Assessment & Evaluation in Higher Education*, 34(3), 305-319.
- Bull, J., Conole, G., Danson, M., Davis, H., Sclater, N., & White, S. (2002). Rethinking sassessment through learning technologies. 19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education, Auckland, New Zealand. 75-86.
- Bull, J., & McKenna, C. (2004). *Blueprint for computer-assisted assessment*. London: RoutledgeFalmer.

- Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *Association for Learning Technology Journal (ALT-J)*, 13(1), 17-31.
- Dearnley, C., Taylor, J., Hennessy, S., Parks, M., Coates, C., Haigh, J., et al. (2009). Using mobile technologies for assessment and learning in practice settings: Outcomes of five case studies. *International Journal on E-Learning*, 8(2), 193-207.
- de-Freitas, S., & Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Computers & Education*, 46(3), 249-264.
- de-Marcos, L., Hilera, J. R., Barchino, R., Jiménez, L., Martínez, J. J., Gutiérrez, J. A., et al. (2010). An experiment for improving students performance in secondary and tertiary education by means of m-learning auto-assessment. *Computers & Education*, 55(3), 1069-1079.
- Dillenbourg, P., Schneider, D., & Synteta, P. (2002). Virtual learning environments. In *Proceedings of the 3rd congress on Information and Communication Technologies in Education*, Rhodes, Kastaniotis Editions, Greece, 3-18.
- Dougiamas, M., & Taylor, P. C. (2003). Moodle: Using learning communities to create an open source course management system. Lassner & C. McNaught (Eds.), *ED-MEDIA2003: Proceedings of the 15th World Conference on Educational Multimedia and Hypermedia & World Conference on Educational Telecommunications*, Honolulu, Hawaii, USA. 171-178.
- Elliott, B. M. (2008). *Modernising assessment in the age of web 2.0*. Scottish Qualifications Authority.
- FREMA. (2007). Reference model for the assessment domain. Retrieved March, 2012, from <http://wiki.frema.ecs.soton.ac.uk/uploads/3/3c/EntityConceptAlphabetic.JPG>
- Google (2010). Google Maps API. Retrieved March, 2012, from <http://code.google.com/intl/es-ES/apis/maps/index.html>
- Google (2011). Google body browser. Retrieved March, 2012, from <http://bodybrowser.googlelabs.com/>
- Harchay, A., Cheniti, L., & Braham, R. (2010). An investigation of the enhancement and the formal description of IMS/QTI specification for programming courses. *International Conference on Advanced Learning Technologies*, Sousse, Tunisia. 113-115.
- Herbert, T. (2010). Learning spaces, learning environments and the dis‘placement’ of learning. *British Journal of Educational Technology (BJET)*, 41(3), 502-511.
- HotPotatoes. (2011). Hot potatoes. Retrieved March, 2012, from <http://hotpot.uvic.ca/index.php>
- Hwang, G., & Chang, H. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education*, 56, 1023-1031.
- Ibáñez, M. B., Morillo, D., Santos, P., Pérez, D., García, J. L., Hernández-Leo, D. & Delgado, C. (2011). Computer Assisted Assessment within 3D Virtual World. *Proceedings of the 1rst European Immersive Education (E-iED)*, 2011, Madrid, Spain. 55-65
- ICOPER. (2011). Interoperable content for performance in a competency-driven society. Retrieved March, 2012, from <http://www.icoper.org/>
- IMS. (2006). IMS question & test interoperability specification v2.0/v2.1. Retrieved March, 2012, from <http://www.imsglobal.org/question/index.html>
- IMS QTI. (2006). QTI v2.1 information model. Retrieved March, 2012, from [http://www.imsglobal.org/question/qti\\_v2p1pd2/imsqti\\_info\\_v2p1pd2.html](http://www.imsglobal.org/question/qti_v2p1pd2/imsqti_info_v2p1pd2.html)
- JISC. (2010). *Effective assessment in a digital age*. Bristol, UK: Retrieved March, 2012, from <http://www.jisc.ac.uk/digiassess>
- Jong, M. S. Y., Lee, F. L., Lee, J. H. M., & Shang, J. J. (2010). A case study of a non-gamer Student’s learning process in VISOLE. *Proceedings of the 3rd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL)*, Kaohsiung, Taiwan. 77-84.
- Joosten, D., Van-Bruggen, J., Hermans, H., Latour, I., Burgers, J., Giesbers, B., et al. (2007). Modeling assessment for re-use of traditional and new types of assessment. *Computers in Human Behavior*, 23(6), 2721-2741.
- Kleppe, A., Warmer, J., & Bast, W. (2003). *MDA explained: The model driven architecture: Practice and promise*. Inc. Boston, MA, USA: Addison-Wesley Longman Publishing Co.
- Koch, D. A. (1993). Testing goes graphical. *Journal of Interactive Instruction Development*, 5, 14-21.
- Liu, G., & Hwang, G. (2010). A key step to understanding paradigm shifts in e-learning: Towards context-aware ubiquitous learning. *British Journal of Educational Technology*, 41(2), E1-E9.
- Luecht, R. M., & Clauser, B. E. (2002). Test models for complex computer-based testing. In C. N. Mills, M. T. Potenza, J. J. Fremer, & W. C. Ward (Eds.), *Computer-based testing: Building the foundation for future assessments*. Mahwah, NJ: Lawrence Erlbaum Associates. 67-88.

- Macdonald, J. (2003). Assessing online collaborative learning: Process and product. *Computers & Education*, 40(4), 377-391.
- Mayotte, S. (2010). Online assessment of problem solving skills. *Computers in Human Behavior*, 26(6), 1253-1258.
- McBeath, R. J. (Ed.). (1992). *Instructing and evaluating higher education: A guidebook for planning LearningOutcomes*. New Jersey: ETP.
- Moodle (2010). Editing a quiz. Retrieved March, 2012, from [http://docs.moodle.org/en/Editing\\_a\\_quiz](http://docs.moodle.org/en/Editing_a_quiz)
- Moodle (2011). Moodle XML format. Retrieved July, 2011, from [http://docs.moodle.org/en/Moodle\\_XML](http://docs.moodle.org/en/Moodle_XML)
- Motiwalla, L. (2007). Mobile learning: A framework and evaluation. *Computers & Education*, 49, 581-596.
- Naismith, L., Lonsdale, P., Vavoula, G., & Sharples, M. (2004). Literature review in mobile technologies and learning No. 11). University of Birmingham: Futurelab.
- Navarrete, T., Santos, P., Hernández-Leo, D., & Blat, J. (2011). QTIMaps: A Model to enable web maps in assessment. *Educational Technology & Society*, 14 (3), 203-217
- Oblinger, D. G. (Ed.). (2006). *Learning spaces*. USA: EDUCASE.
- Open Wonderland Foundation. (2010). Open Wonderland. Retrieved January, 2012 from: Foundation: <http://www.openwonderland.org/>
- Parshall, C., Davey, T., & Pashley, P. (2000). Innovative item types for computerized testing. In Springer (Ed.), *Computerized adaptive testing: Theory and practice*. Netherlands. 129-148.
- Pérez-Sanagustín, M., Hernández-Leo, D., Nieves, R., & Blat, J. (2010). Representing the spaces when planning learning flows. *European Conference on Technology Enhanced Learning (EC-TEL)*, 276-291.
- ReCourse. (2009). ReCourse: LdAuthor. Retrieved March, 2012, from <http://www.tencompetence.org/ldauthor/>
- Ridgway, J., & McCusker, S. (2003). Using computers to assess new educational goals. *Assessment in Education*, 10(3), 309-328.
- Santos, P., Pérez-Sanagustín, M., Hernández-Leo, D. & Blat, J. (2011). QuesTInSitu: From tests to routes for assessment in situ activities. *Computers & Education*, 57 (4), 2517-2534
- Shang, J. J., Jong, M. S. Y., Lee, F. L., & Lee, J. H. M. (2008). VISOLE: An example of hybrid learning I. In J. Fong, R. Kwan, & F. Wang (Eds.), *Hybrid learning and education: Lecture notes in computer science*. Heidelberg: Springer Berlin. 348-358.
- SIMILE. (2003). SIMILE project: TimeLine. Retrieved March, 2012, from <http://simile.mit.edu/timeline>
- Spanish government. (2006). *Real Decree 1631/2006. Department of Education and Science (In Spanish)*. Retrieved March, 2012, from <http://www.boe.es/boe/dias/2007/01/05/pdfs/A00677-00773.pdf>
- Thompson, B. (2005). The transforming effect of handheld computers on nursing practice. *Nursing Administration Quarterly*, 29(4), 308-324.
- Triantafillou, E., Georgiadou, E., & Economides, A. A. (2008). The design and evaluation of a computerized adaptive test on mobile devices. *Computers & Education*, 50(4), 1319-1330.
- Wills, G., Bailey, C., Davis, H., Gilbert, L., Howard, Y., Jeyes, S., et al. (2007). An e-learning framework for assessment (FREMA). In: *International CAA Conference*, Loughborough, UK.
- Zenisky, A. L., & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15, 337-362.
- Zhang, P., Wills, G., Gilbert, L., Warburton, B., & Bacigalupo, D. (2010). IMS QTI engine on android to support offline mobile learning. Paper presented at the *Proceedings of the Computer Assisted Assessment Conference*, Southampton, U.K.