IMS-QTI sub-standards in computerised adaptive testing and interfacing

Martin Lesage*, Martin Riopel, Gilles Raîche and Komi Sodoke

Département d'éducation et pédagogie
Université du Québec à Montréal,
C.P. 8888, succ. Centre-ville,
Montréal(Qc), Canada, H3C 3P8
E-mail: lesage.martin.3@courrier.uqam.ca
E-mail: riopel.martin@uqam.ca
E-mail: raiche.gilles@uqam.ca
E-mail: sodoke.komi@courrier.uqam.ca
*Corresponding author

Abstract: In recent years, a growing number of schools, colleges, universities and other learning institutions have been converting their existing courses into E-Learning applications. To implement this, the authors are developing a Macromedia Flash E-Learning Web application able to fully include item data input and adaptive testing capabilities using item response theory. The project's goal is to maximize the capabilities and the reusability of the multimedia content produced by using the IMS-QTI standard. The application's adaptive testing functionalities will be implemented by proposing new IMS-QTI sub-standards for item parameters and interfacing parameters' characterization.

Keywords: E-Learning; assessment tasks; computerized adaptive testing; computerized assisted learning; learning objects; interface standards; XML.

Reference to this paper should be made as follows: Lesage, M., Riopel, M., Raîche, G. and Sodoke, K. (2008) 'IMS-QTI Sub-standards in Computerized Adaptive Testing (CAT),' *Int. J. Advanced Media and Communication*, Vol. ?, No. ?, pp. xxx–xxx.

Biographical notes: Martin Lesage is a Ph.D. student in the Education Faculty at the Université du Québec à Montréal. He received his master's degree in information technology from the Université du Québec à Montréal and his bachelor's degree in computer science from Université Laval in Québec City. The current paper is part of his Ph.D. research. Martin Lesage is also a sergeant in military engineering for the Canadian reserve forces where he has served as an instructor in many courses related to military engineering. He worked as a multimedia programmer for the Distributed Training Cell of the Canadian Forces Base in Gagetown, New Brunswick for four years. His main research subjects are E-learning standards.

Martin Riopel is a professor in the Education Faculty at the Université du Québec à Montréal. He received his Ph.D. in science education and his master's degrees in education and astrophysics from the Université de Montréal. He is a member of the interuniversity research group on educational assessment and measurement using information and communications technologies (GRIEMEtic) of the Université de Montréal and the Item Response Theory Applications Centre (CAMRI) of UQAM. He is also the founder of the MECAS laboratory for virtual science education. His main research subjects are science education and E-learning.

Gilles Raîche is a professor in the Education Faculty at the Université du Québec à Montréal. He received his Ph.D. in measurement and assessment from the Université de Montréal, his master's degree in education from the Université de Sherbrooke and his bachelor's degree in psychology from Université du Québec à Trois-Rivières. He is a member of the interuniversity research group on educational assessment and measurement using information and communications technologies (GRIEMEtic) of the Université de Montréal and the Item Response Theory Applications Centre (CAMRI) of UQAM. His main research subjects are evaluation, assessment, computerized adaptive testing and E-learning.

Komi Sodoke is a Ph.D. student in the Education Faculty at the Université du Québec à Montréal. He received his master's degree in computer science from the Université du Québec à Montréal and his bachelor's degree in computer science from the University of Moncton. The current paper is part of his Ph.D. research. His main research subject is adaptive testing.

1 Introduction

This research paper concerns assessment and measurement fields in the education domain. It presents a research project aimed at developing computerized distance learning software with formative and summative evaluation capabilities to be distributed over the Internet ("E-Learning applications"). Its main goal is to include adaptive testing in evaluation tasks by creating computer standards that could model both the item parameters of computerized adaptive testing and the interface parameters precisely controlling the display of evaluation tasks.

The project considers the following four aspects of the learning field: evaluation, individualization, adaptation and standardization. It also touches some computer science aspects: computerized adaptive testing, networking with item response theory, self-configurable graphics user interface modeling, XML standards improvement and E-learning application development.

From a technical perspective, the project will use results from the adaptive testing field, some E-Learning standards, the Macromedia Flash ActionScript programming language and theories about user interface ergonomics. It will then propose a new implementation of IMS-QTI sub-standards designed to include XML testing objects that will be parsed by the Flash user interface. Finally, it will propose a self-configurable E-Learning application with adaptive functionalities supported by evaluation objectives based on a new standard that will facilitate their distribution over the Internet.

This paper will first present a theoretical background of adaptive testing, and state user interface ergonomics of E-Learning applications, and invoke a constructivist paradigm for E-Learning and E-Learning standards. Finally, it will present the current results and discuss future developments of the project.

2 Background

The essential notions of item response theory, computerized adaptive testing and implementation of learning methods using item response theory are presented in Wainer (2000). Item Response Theory (IRT) can be summarized as a collection of mathematical models allowing for a formal representation of assessment item characteristics, using statistics to estimate the ability level of a student answering a test (Sodoke, Nkambou, Raîche, Riopel and Lesage, 2007). The three canonical item parameters are discrimination (*a*), difficulty (*b*) and pseudo-guessing (*c*) parameters. The conditional probability for a person with an ability θ to get a correct response to an item *i* with a_i, b_i and c_i parameters is given by:

$$p_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-Da_i(\theta - b_i)}},$$

where *D* has a constant value of 1.701 , so the probability distribution of θ approximate a standardized normal one N(0,1).

IRT provides strategies for:

-estimating learner ability θ (Baker, F. 2001) and its standard error (Sθ);

-estimating item parameters from data (Baker, F. 2004);

-ascertaining how well data fits a model, for instance the Lz misfit index; and

-investigating psychometric properties of assessments.

All these notions are also discussed by Raîche (2004) in relation to the Canadian province of Québec's introduction of computerized adaptive learning applications with the following applications: CAPT (Raîche, 2000, ch. 9, p. 6), FrenchCapt and SIMCAT (Raiche*, 2000, ch. 9, p. 30).*

The constructivist approach of distance learning and some Web based learning models are proposed by Bonk & Wisher (2000). These authors' principal interest is the development of E-learning applications for the military. User interface customization and adaptability have been studied and implemented by Weld, Anderson, Domingos, Etzioni, Gajos, Lau & Wolfman (2004).

The consortium in charge of the IMS-QTI E-learning standard, IMS Global Learning Consortium, describes its XML syntax (www.w3.org/XML/) based E-Learning standards and learning methods implementation on its Web site (www.imsglobal.org). This reference also describes IMS-QTI learning and testing objects.

Recent years have seen the development of applications similar to our project, client-server E-Learning applications with question data entry interfaces and adaptive testing based question (item) display engines where items are stored in a database on a Web server. We will mention four applications in our field of development: SIETTE (Conejo, Guzmàn, Millàn, Trella, Pérez-De-La-Cruz, J.L., & Rìos, 2004), QTIeditor (Pacurar, Trigano & Alupoaie, 2005), CosyQTI (Lalos, Retalis & Psaromiligkos, 2005) and PersonFIT (Sodoke, Kkambou, Raîche, Riopel & Lesage, 2007). The last three use the IMS-QTI standard for question encoding. Another interesting application of E-Learning is the RATH (Relational Adaptive Tutoring) system using the knowledge space theory (Hockemeyer & Albert, 1999; Hockemeyer, Held & Albert, 1997). SIETTE and PersonFIT applications use an evaluation process based on item response theory but do not have self-configurable user interfaces.

A brief introduction to E-Learning standards can be found in Michel & Rouissi (2003) and also in Dunand, Fernandes & Spang-Bovey (2006). Learning objectives and their relation to instructional design theory are defined in Wiley (2000).

A constructivist distance learning approach states that distance learning applications should include evaluation functionalities that allow students to assess their knowledge in a formative evaluation context (Bonk & Wisher, 2000). Distance learning implementations in the academic curriculum should be one element of learning methodology and should consider distance learning application's potential to adapt to student knowledge level, personality and grade.

Many formal E-learning standards are now in use: Dublin Core Metadata Initiative (http://www.dublincore.org), IEEE LTSC LOM (http://ieeeltsc.org), IMS-QTI (http://www.imsglobal.org), AICC/CMI (http://www.aicc.org) and ADL/SCORM (http://www.adlnet.org). Dublin Core, SCORM, LOM and AICC/CMI E-Learning standards do not have the item modeling capacities for evaluation that IMS-QTI has. Nonetheless, while the IMS-QTI standard is able to model some types of assessment items, it still presents two main weaknesses for which a solution is proposed in this paper. The first weakness of the IMS-QTI standard is that it does not model item parameters according to different item response theory models. The second weakness of the IMS-QTI standard is that it does not model user interface display parameters such as backgrounds, buttons, menus, font type, etc. Despite these weaknesses, we believe that the IMS-QTI standard is the most appropriate one designed so far to consider assessment tasks and this is why we are proposing an IMS-QTI sub-standard in this paper.

Related work has been done by Gerbé, Raynauld & Beaulieu, M. (2006) in a project called "Sac d'école électonique" (electronic schoolbag) developed at the Maison des technologies de formation et d'apprentissage Roland-Giguère (MATI) (http://www.matimtl.ca). The goal of the project was to define an XML E-learning standard for learning and evaluation situations in the context of the competency based approach. The Gerbé, Raynauld & Beaulieu project models learning and evaluation situations according to the LOM based Normetic (http://www.normetic.org) standard, instead of the IMS-QTI standard. The project models learning and evaluation situation parameters as general attributes of the situation (title, author, abstract, standard to attain and grade), disciplines (literature, science, etc.), training subjects (academic, entrepreneurship, etc.), development bases (homework, project, etc.), concepts, competencies, evaluation criteria, teaching techniques, and the learning activities of the situation. Even if the LOM based Normetic standard approach has good general classification functionalities, it fails to take account of the quantitative parameters of computerized adaptive testing and item response theory with sufficient precision.

The inclusion of adaptive testing in evaluation tasks will enable teachers to better evaluate their students' abilities. The introduction of specific interface parameters will facilitate the adaptation of the distributed content to the age of the students, their prior knowledge of the subject, the cultural context, etc.

3 Research topic

The rapid growth of existing course material conversion to an electronic format is now an inexorable trend. The research topic addressed in this project is: "How to implement E-Learning in the academic curriculum." Many courses are now converted to electronic format but some problems arise in conversion standardization. One such challenge is the difficulty of exchanging electronic content. User interfaces generally display course content or assessment tasks in a rigid format during an entire presentation sequence of the lesson, exam or work session. These user interfaces generally do not adapt to multimedia constraints, resources or realistic limitations, as required for an optimal presentation of course material or exam questions.

Many distance learning applications distributed over the Internet (E-learning applications) display course material without necessarily including assessment tasks or activities to support learning with formative or summative evaluation. This research project is intended to improve the assessment or evaluation capabilities of E-learning applications. A general approach to the issue will produce answers to the following question: "How do we implement distance evaluation in the academic curriculum?"

Adaptive testing implementation will be accomplished with two IMS-QTI sub-standards; the first sub-standard will be designed to model the item parameters according to the item response theory associated with computerized adaptive testing. The second sub-standard will model user interface parameters such as screen size, backgrounds, menu size and disposition, button size and disposition, etc. A user interface using these standards will be able to adapt to different kinds of assessment tasks to be administered, as well as to different kinds of learners (students, military and workers). Evaluation functionalities will be improved by the addition of such multimedia content as movies, images, video clips and graphic animation to the assessment task for a more realistic contextual approach to evaluation.

The constructivist approach to distance learning states that E-Learning applications should include evaluation capabilities to allow students an opportunity to test their knowledge in a formative evaluation context (Bonk & Wisher, 2000). This project also situates itself in a learning constructivist approach, with a vision of knowledge construction using contextual learning during the production of evaluation tasks in an authentic context. The project's interdisciplinary aspect includes networking and computer science disciplines. The networking aspect encompasses the development of databases and servers storing XML based question banks on the Internet. The computer science aspect covers programming for user interface development and formalization of XML based IMS-QTI sub-standards.

The implementation of E-Learning in the academic curriculum should be a part of learning methodology and consider the adaptation of the application to the student's personality and grade. The current project is intended to improve the process through proposing learning methods supporting an E-Learning application with evaluation and interface self-configuration capabilities based on the IMS-QTI E-Learning standard. This project takes account of the multidisciplinary fields of education in bringing together adaptive testing, learning methods implementation, computer programming of E-Learning applications and networking, by launching and storing these applications over the World Wide Web.

4 Theoretical framework

4.1 Adaptive testing using item response theory

The adaptive testing process could be done without a computer by a human examiner choosing questions to be answered by the examinee (the student) according to the answers given by the examinee to the selected questions. Computerized adaptive testing follows a similar process but the difference is that the examiner is replaced by a computer that processes the answer given by the examinee to choose the next question according to statistical estimators selected by item response theory.

A formal definition of adaptive testing could be a test question selection and display process in which test questions named "items" are selected one by one from a question bank to be shown to the student according to the validity of the answers of the previously administered questions. If the answers given by the student are mostly incorrect, easier questions will be selected. In the opposite case, if the answers are mostly correct, harder questions will be selected from the question database, as shown in Figures 1 and 2.

The question selection and administration process adapts themselves to the examinee by tailoring a unique exam to a specific student, giving rise to the term “adaptive testing” as shown in Table 1 (Weiss & Kingsbury, 1984, p. 361; Stocking, 1996, p. 4; Raîche, 2000, p. 7).

Figure 1 Adaptive testing algorithm using item response theory (adapted from Wainer, 2000, p. 106 and Raîche, 2004)

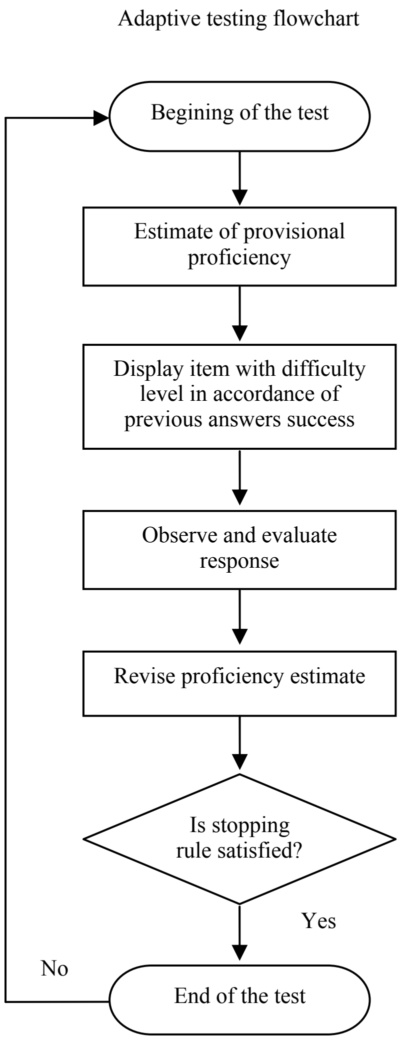
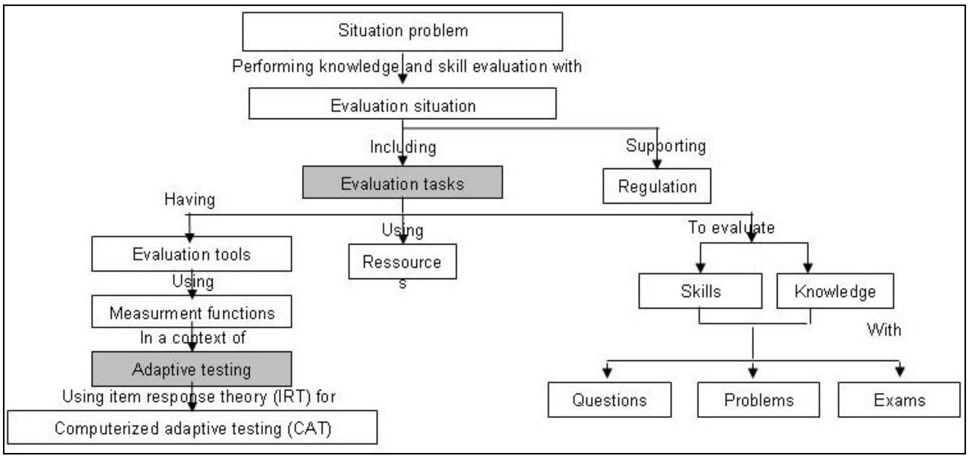


Figure 2. Situation-problems' hierarchy of concepts



The "item" term of the expression "item response theory (IRT)" refers to a test or an exam question answered either with a written answer in a traditional manner or a multiple choice selection. The answer must be easy to convert into electronic format to be inputted into a computerized system or converted into an application in the case of computerized adaptive testing (Auger, 1989; Raîche, 2000).

Table 1 Adaptive testing rules based on item response theory (Raîche, 2004)

ALGORITHM OF ADAPTIVE TEST BASED ON ITEM RESPONSE THEORY

Rule	Action
Starting instruction	Submit an item (a question of the test) with difficulty level in accordance with the student profile (personality, knowledge level, etc.).
Looping condition	Submit an item with difficulty level near the skill level provisional proficiency estimate.
Stopping rule	Stop the test: When some predetermined number of items has been administered Or: When the skill level provisional proficiency estimate indicates a predetermined type of error/There is an occurrence of skill level provisional proficiency estimate predetermined type of error Or: When no further items could influence the estimation of the student skill level

The use of computerised adaptive testing in the field of education refers to an approach that is different from item response modeling's fundamental methods based on classical testing theory that only considers the number of right answers. Computer usage in the education field is expanding rapidly because these data processing machines have become more affordable. The use of computers increases the processing capabilities of the educational field researchers, especially for mathematical and statistical data processing. New statistical methods to evaluate student abilities to answer test questions, stemming from "latent trait theory," now constitute the basis of item response theory, as used in computerized adaptive testing (Raîche, 2000, p.7). A formal definition of adaptive testing is given in Auger's Ph.D. thesis:

> "adaptive testing defines itself as an item selection strategy resulting in the administration of items that only allows a fair measure of its abilities: every success is followed by a more difficult item and each failure by an easier item, selected into the non-answered items." (AQuger, 1989, p. 17)

The question selection process mentioned in Auger's definition is based on an estimate of the student's ability level. This estimation stems from latent trait theory, based on mathematical models and statistical estimators that analyse variables defined by item response data analysis to determine ability level and students' probability of answering test questions (items) correctly (Weiss, 1983, p. 1; Eggen & Straetmans, 2000, p. 713).

As stated previously, adaptive testing can be done in a non-computerized manner. Some adaptive testing processing methods are designed to avoid latent trait computation by splitting the question bank into two types of items: items to be presented in case of a correct answer by the student and other items to be presented in case of incorrect answers. The insertion of item response theory in adaptive testing is now possible because of computer capacity to process information on answers in testing sessions. The use of the computer and item response theory allow statistical estimation of latent traits through processing all answers to questions previously shown during the test. When the computer is used in adaptive testing, we use the expression "computer adaptive testing (CAT)" (Weiss & Kingsbury, 1984, p. 361).

The processing of a computer adaptive test has two phases. The first phase is the statistical estimation of student abilities, considering their answers to the questions. The second phase is the processing of answers and statistical estimations for the choice of the next question. The computer adaptive test sequence proceeds as follows: the student is shown a question, the student answers the question, the response is entered and processed by the computer system and estimators for the selection of the next question are computed. The next question is now retrieved from the item database according to the estimators and presented to the student. The administration of test questions ends when the ability of the student reach a predetermined precision criteria (standard error usually). The test could also end when the student has answered a predetermined number of questions or when all the items in the question bank have been presented, as shown in Figure 2 (Weiss & Kingsbury, 1984, p. 361; Eggen & Straetmans, 2000, p. 713).

Adaptive testing theory adds flexibility to testing by selecting the questions in accordance with the user's knowledge level, as shown in Figure 2. Once the first question is selected in accordance with the student's skills, subsequent questions are selected depending on the student's answers to the previous questions.

Auger stated that this evaluation method increases measurement precision:

> "… item selection is repeated until the ability level of the individual could be estimated with a predetermined of accuracy or until a maximal required number of items or all of the database items were presented. In theory, this proposed strategy reduces testing duration and the number of administered items, and increases measurement precision." (Auger,1989, p. 17)

Computerised adaptive tests were initially developed to compute accurate estimations of students' abilities to attain objectives, and their abilities or level of knowledge in terms of grades or graduating. Research and improvements in the computerized adaptive testing field could now have purely formative objectives, allowing students to test their skills or knowledge, without being graded. Computerized adaptive testing also classifies students' abilities or knowledge level (beginner, average, skilled, or advanced) to place them in a particular academic curriculum (Eggen & Straetmans, 2000, p. 713).

Wainer & al. (2000) states some advantages of adaptive testing over standard evaluation which are security concerns about the possibility to decrement plagiarism and permits learning flexibility allowing the student to interact at his own pace. The main goal of computerized adaptive testing is to challenge students instead of stressing or discouraging them. The students have also some feedback and an instant knowledge of their results or progress. In the computerized adaptive test administration process development, a wider variety of questions could be added in a question bank and non discriminating questions could also be removed from the bank. The test results are instantly entered in a database without some conversion of the student's answers.

A disadvantage of usual adaptive testing strategies compared to traditional exams in printed format is the impossibility of controlling the order of presented and answered questions, making it impossible to have the same test or question sequence for all students. This particularity occurs when, although two students have the same number of right answers, the correctly answered questions differ. This is also true for the sequence of incorrect answers. The dynamic nature of computerized adaptive testing, selecting subsequent questions according to previous answers, means that the order of questions may vary for each student repeating the same test.

4.1.1. Definition of adaptive testing

We will now provide a formal definition of adaptive testing that includes computerized adaptive testing by extension. Due to its algorithmic nature, there is no contradiction between the definitions of the aforementioned authors who are building upon earlier results.

We can define adaptive testing as an application of learning evaluation consisting of the selection and presentation of questions contained in a question bank according to their level of difficulty and student answers to these questions (Weiss & Kingsbury, 1984, p. 361; Stocking, 1996, p. 4). Every question correctly answered is followed by a more difficult question and every question wrongly answered is followed by an easier question. Adaptive testing could serve as formative evaluation (with ungraded goals) and incorporate student classification criteria (Auger 1989, p. 17; Stocking, 1996, p. 4; Eggen & Stratemans, 2000, p. 713).

One special field of adaptive testing is computerized adaptive testing. It is an electronic process of question selection computerized process using a database. The subsequent question is determined by item response theory statistical estimators (Weiss & Kingsbury, 1984, p. 361; Eggen & Straetmens, 2000, p. 713).

The sequence of a computerized adaptive test is described by the following algorithm: the first question is either randomly selected or chosen according to an estimation of the student's initial level of knowledge. The following questions are selected depending on the previous answers, using item response theory modelization. The test sequence ends when a predetermined number of questions are presented or when level of knowledge of the student is computed with a predetermined accuracy (Weiss & Kingsbury, 1984, p. 362; Eggen & Straetmans, 2000, p. 713).

4.2 Learning objects

Learning objects are sharable and reusable electronic entities (data structures) used for learning. These entities are always available at any time and not stored only on the Internet. They are usually stored in data repositories controlled by their administrators or developers. Learning objects can be entire courses or component elements such as audio/video clips, images or graphic animation.

4.3 Testing objects or evaluation objects (Items)

An item is a set of digitized interactions (that could be void),, with all the attendant support material, that can be analyzed with a set of rules permitting the conversion of a candidate's answer into evaluation results. Item size could vary from a single question composed of text and one input field to an entire multiple question exam with instructions and multimedia support material (Michel & Rouissi, 2003).

4.4 Evaluation tasks

An evaluation task could be any kind of task including questions or problems used by teachers, and could be presented independently of objectives or competency learning approaches, as stated by Tousignant:

> "When we submit a test or an exam, we ask the student to say or do something to show us what he learned, what he understood and the level of mastery of the skills we want the student to develop. We ask him questions needing an oral or written answer; we submit a problem to him to which he must show the solution; we invite him to perform a task that we could further analyze." (Tousignant,1982, p. 16)

Evaluation tasks are complex situation-problems used in a precise and authentic context. They are powerful tools to develop skills through evaluation situations, using diverse resources that could include knowledge, skills, abilities and attitudes belonging to many disciplines in order to perform in an authentic, realistic context.

Evaluation tasks are used to assess student knowledge in various disciplines, such as English, mathematics, literature, chemistry, etc. Evaluation tasks are part of the education evaluation and measurement field. They could be simply exams or more complex assessment processes like text analysis for orthography and grammar evaluation. They can also be used to assess skills or abilities, like the ability to perform certain sports or games in physical training courses (Scallon, 2004, p. 113).

Evaluation tasks also include common sense skills assessment of student effectiveness in time and schedule planning, the ability to work as part of a team, the ability to research information and problem-solving capabilities.

As shown in Figure 2, situation-problems could include evaluation tasks. We will now define the situation-problems concept. Scallon (2004, p. 112) describes the situation problem concept as "all kinds of complex tasks, all projects challenging the student by using resources" and adds that it is "a generic term that could not include problem solving in the strictest sense." He also states that the expression "situation-problems" comes from mathematics instruction where students were provided written problems describing a situation requiring the application of mathematical formulae and numerical computation.

As noted in previous sections, the framework of the situation-problems context encompasses all tasks calling for student resource utilization. The situation-problems definition does not specify that they could include evaluation tasks. To eliminate any confusion created by previous definitions, we will consider that evaluation tasks are distinct from situation-problems as shown in Table 2. According to the literature, if situation-problems need to have an evaluation session, it will be included in an evaluation situation that will be part of the situation-problems, as shown in Figure 2.

Table 2 Evaluation tasks description

Type	Definition	Examples
Situation-problems	All kinds of complex tasks, all projects challenging the student by requiring the use of resources Do not need to include questions or problems to solve Could include evaluation situations but do not need to include evaluation defined by criteria	Arts and crafts Textbook reading Written or oral exam Games Watching television
Evaluation situation	Complex tasks generally designed by governments or departments of education to assess a broad population of students in terms of particular skills. Includes measurement functions with evaluation tasks Evaluation criteria are defined by governments or education departments	Admission tests Education departments' exams Physical training tests Building models Scientific observations Mathematical problem solving
Evaluation tasks	All types of tasks including questions or problems They are included in the education field of measurement and evaluation of students' knowledge and skills assessment in multiple disciplines Evaluation criteria are defined by the teacher or by the academic institution	Exams Common sense knowledge Skill or ability evaluation, such as the capacity to play a particular sport

Other authors such as Durand and Chouinard (2006, p. 129) also agree with these definitions and describe evaluation tasks as part of evaluation situations. According to these authors, evaluation situations are complex tasks designed by governments or departments of education to assess a particular skill for a wide population of students. Evaluation situations are also part of the regulation process of the education system with respect to skill development. The nature of the problem described in evaluation situations brings together all the objectives of the skill. They also include formal evaluation tools and results, as well as interpretation methods.

The evaluation situation arises from an analytical approach. The approach is based on criteria evaluating students' abilities to solve complex problems using resources. The evaluation situation can provide information on skills and knowledge levels attained by the student. It also offers support for the regulatory process, especially retroactive regulation, following a learning activity (Boucher, Loiselle & Reiber, 2006, p. 3). The teacher's regulation process is intended to inform students about their acquired knowledge or abilities and about those they need to improve (Durand & Chouinard, 2006, p. 88).

Louis and Bernard (2004, p. 79) state that complex learning task evaluation "is based on the presentation of tasks oriented to the integration of the acquired skills or knowledge by the student" and that "the evaluation task in an authentic context is designed to measure a set of affective and cognitive dimensions allowing for its more effective realization."

Durand & Chouinard (2006, p. 72) conclude this section with a list of examples of complex tasks, such as the building of a flying object mock-up or the organization of an election in ancient Greece or any other task using an authentic context, such as mathematical problem solving, scientific observations or laboratory manipulations.

4.4.1. Evaluation tasks' definition

The previous sections allowed us to distinguish amongst a myriad of expressions related to knowledge construction, such as "situation-problems," "evaluation situations" and evaluation tasks. As shown in Figure 2, situation-problems include evaluation situations that could have one or many evaluation tasks.

Scallon (2002) states that situation-problems include every complex project allowing the student to draw upon resources. The expression "situation problem" comes from written problems in the field of mathematics

Evaluation situations include evaluation tasks and ask students to solve complex problems (Durand & Chouinard, 2006). They can also allow assessment of students' knowledge and skill level and support the retroactive regulation process (Boucher, Loiselle & Reiber, 2006, p.3).

The preceding sections allow us to define evaluation tasks synthesizing previous definitions: an evaluation task could be all types of tasks involving questions or problems to analyse (Tousignant, 1982, p. 16). These tasks are generally used to assess various skills or knowledge in many situations (Tardif, 2006, p. 125; Durand & Chouinard, 2006, p. 72; Louis & Bernard, 2004, p. 25).

4.5 Computer standards

A standard in defined by the dictionary of computing (Illingworth, 1996) as: "A publicly available definition of a hardware or software component, resulting from international, national, or industrial agreement" and also "A product, usually hardware, that conforms to such a definition."

The goal of the present research project is to define two formal language XML-based computer standards. The first provides a formal definition of user interface characterization parameters and the second defines item parameters of item response theory. These two computer standards are designed using the XML based E-learning standard IMS-QTI because this standard can model evaluation objects.

4.5.1. The XML computer standard

The XML acronym stands for eXtensible Markup Language. It is a tag-based language similar to HTML Web page definition language. The difference between HTML and XML is that, for XML, the tags are defined by the programmer, unlike HTML which has fixed tags. The XML language is especially designed to format file data to store information on Web server databases. E-learning standards like LOM, SCORM and IMS-QTI are built around this language to format course material data into learning objects and evaluation objects.

4.6 Learning and evaluation situations

The learning and evaluation situations concept resides in the application of theoretical concepts to real situations. They have an evaluation component that is the evaluation situation, to determine whether the student has learned the skills or concepts to solve real life problems. A general definition of the learning and evaluation situation concept is a way to plan teaching to develop skills and knowledge according to the course program. It includes tasks to develop disciplinary and transversal competencies. It includes an evaluation process to continuously promote learning and to evaluate the level of development of competencies (MELS, 2007; Boucher, Loiselle & Reiber, 2006, p.2). The terms “Learning and evaluation situations” come from the old terminology of “learning activities proposition” and “pedagogical scenarios.” Learning and evaluation situations are designed to be more complex and systematic and to always include an evaluation component. They must allow every student to access the resources needed for knowledge construction and to use their skill (Bibeau, 2007, p.1). Learning and evaluation situations have three pedagogical intervention stages, preparation, achievement and integration of learning (Gerbé, Raynault & Beaulieu, 2006, p. 2). They have two principal elements, the topic, often formulated in a question, and a set of tasks and learning activities. They can be built in six steps: subject delimitation, the search for references, reference selection, accurate information selection, information processing and construction of the learning situation (EBSI, 2007). Finally, the learning and evaluation situations include the evaluation situation.

4.6.1. Learning and evaluation situation definition

Learning and evaluation situations include evaluation situations. Their contribution resides in the application of theoretical concepts to real situations. They assess whether the student has acquired the skills or learned the concepts to solve real life problems, according to the course program.

4.7 The IMS-QTI evaluation standard

The IMS-QTI standard allows for the packaging (encapsulation) of course sections or teaching points (“learning objects ”) into small XML (“eXtensible Markup Language”) modules as shown in Figure 3.

Many E-learning formal standards are now in use: Dublin Core/DCMI(http://www.dublincore.org), IEEE LTSC LOM (http://ieeeltsc.org), IMS-QTI (http://www.imsglobal.org), AICC/CMI (http://www.aicc.org), and ADL/SCORM (http://www.adlnet.org)

Dublin Core, SCORM, LOM and CMI standards are especially designed to automate and model course material (learning objects). These standards don’t include formalism and item modeling parameters for testing (learning evaluation).

The IMS-QTI standard does not currently allow online adaptive testing based on item response theory (IRT).

To solve this problem, this project aims to improve the IMS-QTI standard with the formal implementation of two sub-standards: characterization of interfaces and characterization of item parameters according to item response theory.

Figure 3 A multiple choice question coded in the IMS-QTI standard (Lesage, Raîche, Riopel, & Sodoke, 2007)

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Thie example adapted from the PET Handbook, copyright University of Cambridge ESOL
Examinations -->
<assessmentItem xmlns="http://www.imsglobal.org/xsd/imsqti_v2p0"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.imsglobal.org/xsd/imsqti_v2p0 imsqti_v2p0.xsd"
   identifier="choice" title="Unattended Luggage" adaptive="false" timeDependent="false">
   <responseDeclaration identifier="RESPONSE" cardinality="single" baseType="identifier">
      <correctResponse>
         <value>ChoiceA</value>
      </correctResponse>
   </responseDeclaration>
   <outcomeDeclaration identifier="SCORE" cardinality="single" baseType="integer">
      <defaultValue>
         <value>0</value>
      </defaultValue>
   </outcomeDeclaration>
   <itemBody>
      <p>Look at the text in the picture.</p>
      <p>
         <img src="images/sign.png" alt="NEVER LEAVE LUGGAGE UNATTENDED"/>
      </p>
      <choiceInteraction responseIdentifier="RESPONSE" shuffle="false" maxChoices="1">
         <prompt>What does it say?</prompt>
         <simpleChoice identifier="ChoiceA">You must stay with your luggage at all
times.</simpleChoice>
         <simpleChoice identifier="ChoiceB">Do not let someone else look after your
luggage.</simpleChoice>
         <simpleChoice identifier="ChoiceC">Remember your luggage when you
leave.</simpleChoice>
      </choiceInteraction>
   </itemBody>
   <responseProcessing
      template="http://www.imsglobal.org/question/qti_v2p0/rptemplates/match_correct"/>
</assessmentItem>
```

4.8 Objectives

To define this project's main objectives, we have considered the following works. The RATH application focuses on adaptive learning. Other testing applications like QTIeditor, CosyQTI and PersonFIT use the IMS-QTI format. SIETTE and PersonFIT applications have been developed for adaptive testing using item response theory. Finally, the PersonFIT application uses the IMS-QTI standard in the adaptive testing context. Unfortunately, these applications are not part of any academic curriculum and are not the object of learning methods. These applications also lack a user interface parameter characterization for self-adaptability. In this project, we will work to improve learning methods in adaptive testing. We also want to increase user interface adaptability.

The first objective of this project will be the development of the two IMS-QTI sub-standards, the first used for item parameter modeling and the second for user interface characterization parameter modeling. The second objective will be the design of XML based evaluation tasks using the sub-standards. The third objective will be the development of a distance learning application with a built-in question display engine using these standards.

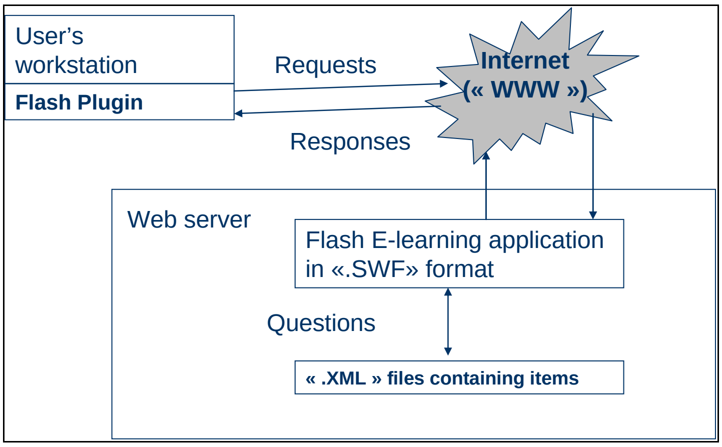
5 Methodology and preliminary results

This project has three phases. The first is the development of an E-Learning application with adaptive testing capabilities, the second is the implementation of QTI-IMS sub-standards for item and user interface parameters modeling, and the final phase is the establishment of learning methods with formative evaluation capabilities.

The E-Learning application will be tested on subjects who are students in Québec schools and members of the Canadian Army. The results will be collected through interviews and adaptive tests using the E-learning application.

The application is based on a client/server architecture. Users access the application with their Web browser (ex.: Internet explorer). The application is hosted on a Web site on a server. It can be used for question data entry and has a question display engine for adaptive testing in accordance with item response theory. The inputted questions are converted into assessment items coded in QTI-IMS XML format and stored in the server database as shown in Figure 4.

Figure 4 Software architecture of CAMRI Laboratory's Flash adaptive testing application for reading and writing XML IMS-QTI standard files on local or user workstation mode. (Lesage, Raîche, Riopel, Sodoke, 2007)



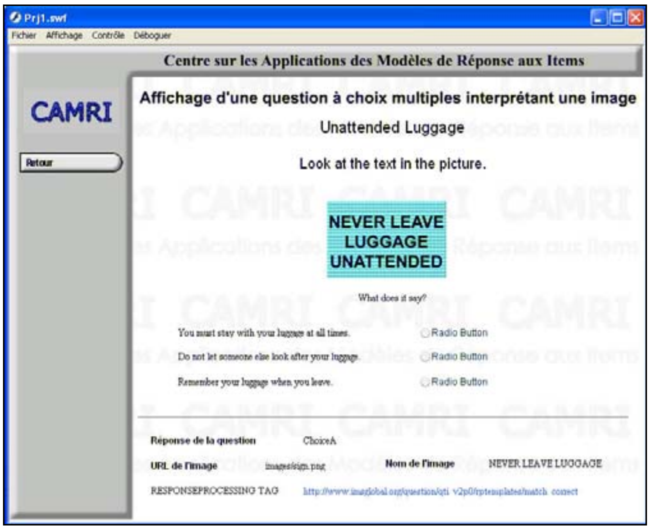
The question data entry interface is shown in Figure 5. The actual stage of development only allows the data entry of multiple choice questions.

The application is able to generate items in IMS-QTI format and to store the items in its database. The application is also able to perform the reverse operation and retrieve items from its database, parse the XML and display the question as shown in Figure 6.

The bulk of the project remains to be done and consists of the completion of the E-Learning application, the implementation of IMS-QTI sub-standards and the accompanying learning methods.

Figure 5 Item (Question) data entry interface of CAMRI Laboratory's Flash adaptive testing application (Lesage, Raîche, Riopel, Sodoke, 2007)

Figure 6 IMS-QTI item displayed in the flash application



5.1 *The IMS-QTI sub-standard for interface parameters*

Our QTI-IMS sub-standard will be similar to the Mozilla XML user interface language (XUL) project (http://www.mozilla.org/projects/xul/). Due to the multiple uses of XML in defining interface and item parameters, the interface parameter characterization sub-standard could not be simply formatted in XML, like XUL, because our standard has to model evaluation objects in IMS-QTI format. We have decided to implement an IMS-QTI sub-standard for interfaces that will be similar to XUL but modeled on the IMS-QTI standard. The main concern is that the interface be able to adjust itself to a sequence of questions in evaluation tasks where different multimedia elements have to be displayed at each question as large images or full screen movies or video clips.

The interface will have to adjust itself to the presentation of a full screen multimedia object through managing its content area and window size. The applications will be able to expand the content area and close some less relevant windows, icons or menus for full screen multimedia presentations.

The standard will also be able to control the display mode and the position of menus, buttons, backgrounds and toolbars. It will also determine all the text in the interface to modify its size, color, font and position.

Some examples of a QTI-IMS sub-standard characterization for interface parameters are shown below:

- Menus
<MenuParameter positionX=“100” positionY=“200” type=“DropDown” />

- Backgrounds
<BackGroundParameter src=“bckdir/bck.png” positionX=“100” positionY=“200” Animated=“No” />

- Buttons
<Button positionX=“100” positionY=“200” type=“Rectangle” caption=“E-mail” Animated=“Yes”>

- Windows
<windows positionLeft=“100” positionTop=“100” positionRight=“600” positionBottom=“700” >

- Video
<video positionLeft=“100” positionTop=“100” positionRight=“600” positionBottom=“700” src= “president.mpeg” >

- Image
<image positionLeft=“100” positionTop=“700” positionRight=“100” positionBottom=“700” src= “img1.jpg” >

- Graphics animations
<graphicsAnimation positionX=“100” positionY=“200” src=“anim/anim1.jpg”/>

5.2 The IMS-QTI sub-standard for item parameters

The project will study the E-learning standard IMS-QTI in detail because it already contains question banks and item modeling parameters. Our goal is to enhance the IMS-QTI standard by developing a sub-standard for item parameter modeling.

As presented before in this paper, the three item parameters associated with item response theory are discrimination (a), difficulty (b) and pseudo-guessing (c) parameters. One must also consider the subject ability represented by θ and the conditional probability associated with item i as represented by *Pi*.

As a numerical example, if Pi = .11, θ = .25, a = .3, b = .4 and c = .5, the new proposed QTI-IMS modeling for item parameters would be:

<ItemParameter P=”0.11” Theta = “0.25 a=”0.3” b=”0.4” c=”0.5”/>

Or when all the parameters are in multiple XML structures:

```
<PItemParameter>0.11</PItemParameter>
<ThetaItemParameter>0.25</ThetaItemParameter>
<AItemParameter>0.3</AItemParameter>
<BItemParameter>0.4</BItemParameter>
<CItemParameter>0.5</CItemParameter>
```

5.3 Inclusion of the IMS-QTI sub standard in XML item definition

Once the sub-standards will be defined, they will be included in the IMS-QTI standard XML formatted items. As stated, the item response theory parameters sub-standards will be included in the items with the interface parameters sub-standards allowing the interface to adapt to the multimedia elements. Figure 7 shows a multiple choice question with an image named “sign.png” in a 400 x 400 pixel windows at 100 pixels from the upper left corner.

Figure 7 A multiple choice question displaying an image (adapted from Lesage, Raîche, Riopel, Sodoke, 2007)

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Thie example adapted from the PET Handbook, copyright University of Cambridge ESOL Examinations
-->
<assessmentItem xmlns="http://www.imsglobal.org/xsd/imsqti_v2p0"
   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
   xsi:schemaLocation="http://www.imsglobal.org/xsd/imsqti_v2p0 imsqti_v2p0.xsd"
   identifier="choice" title="Unattended Luggage" adaptive="false" timeDependent="false">
   <responseDeclaration identifier="RESPONSE" cardinality="single" baseType="identifier">
      <correctResponse>
         <value>ChoiceA</value>
      </correctResponse>
   </responseDeclaration>
   <outcomeDeclaration identifier="SCORE" cardinality="single" baseType="integer">
      <defaultValue>
         <value>0</value>
      </defaultValue>
   </outcomeDeclaration>
  <itemParameter P=”0.15” Theta = “0.25  a=”0.3” b=”0.4” c=”0.5”/>
  <windows positionLeft=“100” positionTop=“500” positionRight=“100” positionBottom=“500” >
   <itemBody>
      <p>Look at the text in the picture.</p>
      <p>
         <img src="images/sign.png" alt="NEVER LEAVE LUGGAGE UNATTENDED"/>
      </p>
      <choiceInteraction responseIdentifier="RESPONSE" shuffle="false" maxChoices="1">
         <prompt>What does it say?</prompt>
         <simpleChoice identifier="ChoiceA">You must stay with your luggage at all times.</simpleChoice>
         <simpleChoice identifier="ChoiceB">Do not let someone else look after your
luggage.</simpleChoice>
         <simpleChoice identifier="ChoiceC">Remember your luggage when you leave.</simpleChoice>
      </choiceInteraction>
   </itemBody>
   <responseProcessing
      template="http://www.imsglobal.org/question/qti_v2p0/rptemplates/match_correct"/>
</assessmentItem>
```

Figure 8 shows the next question displayed after the question in figure 7. The item parameters have changed to reflect the estimators of difficulty for this new question. Instead of a small image, this question is displaying a large 800 x 600 pixels video in a 900 x 1000 window. To control the video, a button assign for a manual start of the video by the student is displayed at coordinates X= 1000 pixels and Y = 1100 pixels.

Figure 8 A multiple choice question displaying a video (adapted from Lesage, Raîche, Riopel, Sodoke, 2007)

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Thie example adapted from the PET Handbook, copyright University of Cambridge ESOL
Examinations -->
<assessmentItem xmlns="http://www.imsglobal.org/xsd/imsqti_v2p0"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="http://www.imsglobal.org/xsd/imsqti_v2p0 imsqti_v2p0.xsd"
    identifier="choice" title="Unattended Luggage" adaptive="false" timeDependent="false">
    <responseDeclaration identifier="RESPONSE" cardinality="single" baseType="identifier">
        <correctResponse>
            <value>ChoiceA</value>
        </correctResponse>
    </responseDeclaration>
    <outcomeDeclaration identifier="SCORE" cardinality="single" baseType="integer">
        <defaultValue>
            <value>0</value>
        </defaultValue>
    </outcomeDeclaration>
   <itemParameter P=”0.45” Theta = “0.65  a=”0.7” b=”0.9” c=”0.33”/>
   <windows positionRight=“50” positionTop=“50” positionLeft=“950” positionBottom=“1050” >
   <button positionX=“1000” positionY=“1100” type=“Rectangle” caption=“Play” Animated=“No”>
<video positionLeft=“50” positionTop=“50” positionRight=“850” positionBottom=“650” src=
“president.mpeg” >
    <itemBody>
        <choiceInteraction responseIdentifier="RESPONSE" shuffle="false" maxChoices="1">
            <prompt>Which president of the United-States is shown in the video?</prompt>
            <simpleChoice identifier="ChoiceA">Lincoln</simpleChoice>
            <simpleChoice identifier="ChoiceB">Carter</simpleChoice>
            <simpleChoice identifier="ChoiceC">Clinton </simpleChoice>
        </choiceInteraction>
    </itemBody>
    <responseProcessing
        template="http://www.imsglobal.org/question/qti_v2p0/rptemplates/match_correct"/>
</assessmentItem>
```

6 Conclusion

The main goal of this research project was to facilitate the implementation of distance learning in education programs by designing computer interface standards that include computerized adaptive testing in evaluation tasks. The fact that the evaluation task also determines the user interface results in a more realistic evaluation context. The standardization of evaluation tasks according to/by XML-based computer standards facilitates the exchange of XML formatted teaching material files (questions, problems, courses and exams) with colleagues in learning institutions worldwide over the Internet without further processing or conversion.

Our decision to standardize evaluation tasks instead of the situation-problems or the evaluation situations means that each new task determines the user interface and changes the menus, buttons, and backgrounds for optimal presentation of the new multimedia elements.

An important aspect of this project will be the implementation of learning methods for E-Learning applications with adaptive testing functionalities in a socio-constructivist formative evaluation context. One of the main goals of these methods is preparing students to use E-testing software to overcome their resistance to change. Another important issue is the summative aspect of the evaluation with a pass/fail criterion.

References

ADL/SCORM (2007). *Sharable Context Object Reference Model (SCORM)* [En ligne]. Accès: http://www.adlnet.org

AICC-CMI (2007). *Aviation Industry CBT Committee (AICC)-Courseware Management Interface (CMI)* [En ligne]. Accès: http://www.aicc.org

Auger, R. (1989). *Étude de praticabilité du testing adaptatif de maîtrise des apprentissages scolaires au Québec. Une expérimentation en éducation économique secondaire 5*. Thèse de doctorat inédite, Université du Québec à Montréal.

Baker, F. B. (2001). *The basics of item response theory*. Dans ERIC clearinghouse on assessment and evaluation (Dir). College Park, MD: University of Maryland [En ligne]. Accès: http://edres.org/irt/baker/

Baker, F. B. (2004). *Item response theory – Parameter estimation techniques*. New York: Marcel Dekker.

Bibeau, R. (2007). Des situations d'apprentissage et d'évaluation (SAE) sur Internet. *Revue de l'association EPI* [En Ligne], 91. Accès : http://www.epi.asso.fr/revue/articles/a0701a.htm

Bonk, C. J. et Wisher, R. A. (2000). *Applying collaborative and e-learning tools to military distance learning: A research framework* (Technical Report 2207). Alexandria, VA: United States Army Research Institute for the Behavioral and Social Sciences.

Boucher, A.C., Loiselle, A.M. et Reiber D. (2006). *Les situations d'apprentissage et d'évaluation… Lexique* [En ligne]. Accès : http://recit.csp.qc.ca/domaine_arts/portfolio/ftp/SAE_lexique_final_mp.pdf

Conejo, R., Guzmàn, E., Millàn, E.,Trella, M., Pérez-De-La-Cruz, J.L. et Rìos, A. (2004). SIETTE : A Web Based Tool for Adaptive Testing. *International Journal of Artificial Intelligence in Education, 14*, 29-61.

Dublin Core Metadata Initiative (2007). *Dublin Core Metadata Initiative* [En ligne]. Disponible: http://dublincore.org

Dunand, N., Fernandes E. et Spang-Bovey, N. (2006). Review: e-learning and standardization. *CETIS briefings on E-Learning standards* [En ligne]. Accès : http://telearn.noe-kaleidoscope.org/warehouse/Fernandes-Emmanuel-2006b.pdf

Durand, M.J. et Chouinard, R. (2006). *L'évaluation des apprentissages*. Montréal : Éditions Hurtubise HMH ltée.

EBSI (2007). *Les 6 étapes d'un projet de recherche d'information* [En ligne]. Accès : http://www.ebsi.umontreal.ca/jetrouve/projet/index.htm

Eggen, T.J.H.M. et Straetmans G.J.J.M. (2000) Computerized adaptive testing for classifying examinees into three categories. *Educational and Psychological Measurement, 60*(5), 713-734.

Gerbé, O. Raynauld, J., et Beaulieu, M. (2006). *Vers un standard de la description des situations d'apprentissage et d'évaluation* [En ligne]. Accès : http://neumann.hec.ca/chaireteag/ColloqueSc%E9nario2007RaynauldFinal.pdf

Hockemeyer, C. et Albert D. (1999). The Adaptive Tutoring System RATH : a Prototype. *Workshop Interactive Computer aided Learning: Tools and Applications, ICL99,* pp. 1-10 [En ligne]. Villach, Austria: Carinthia Tech Institute. Accès: http://wundt.uni-graz.at/rath/publications/icl99.pdf

Hockemeyer, C., Held, T. et Albert D. (1997). RATH A Relational Adaptive Tutoring Hypertext WWW-Environment Based on Knowledge Space Theory. *Proceedings of the Fourth International Conference on Computer Aided Learning in Science and Engineering (CALISCE'98),* 417-423. Göteborg, Sweden: Chalmers University of Technology.

IEEE LTSC LOM (2007). *Learning Object Metadata*. [En ligne] Accès: http://ieeeltsc.org

Illingworth,V. (Ed.). (1996). Dictionary of computing. New York: Oxford University Press.

Lalos, P., Retalis, S. et Psaromiligkos, Y. (2005) *Creating personalised quizzes both to the learner and to the access device characteristics: the Case of CosyQTI* [En ligne]. Accès : http://wwwis.win.tue.nl/~acristea/AAAEH05/papers/3camera%20readyCoSyQTI.pdf

Lesage, M., Raîche, G., Riopel, M. et Sodoke, K. (2007). Interfaces Modelization in Computerized Adaptive Testing (CAT) Context for Learning Optimization. *E-Learn 2007,* 342-351.

Louis, R. et Bernard, H. (2004). *L'évaluation des apprentissages en classe*. Montréal : Groupe Beauchemin, éditeur ltée.

MATI (2008). *Maison des technologies de formation et d'apprentissage Roland-Giguère.* [En ligne] Accès: http://www.matimtl.ca

MELS (2007). *Situations d'apprentissage et d'évaluation* [En ligne]. Québec, Ministère de l'Éducation, du loisir et du sport, Direction générale de la formation des jeunes. Accès : http://www.mels.gouv.qc.ca/dscc/pdf/SAE/SAE.PDF

Michel C. et Rouissi S. (2003). E-learning : normes et spécifications. Étude des spécifications LOM et IMS-QTI caractérisant des documents numériques inter-échangeables et réutilisables pour l'acquisition et l'évaluation des connaissances. *Document Numérique* , *7*(1-2), 157-178.

Pacurar E.G., Trigano P. et Alupoaie S. (2005). A QTI editor integrated into the netUniversité web portal using IMS LD. Dans C. Tattersall et R. Koper (Dirs.) *Journal of Interactive Media in Education: Advances in Learning Design Special Issue* (septembre 2005) [En ligne]. Accès : http://www-jime.open.ac.uk/2005/09/

Raîche, G. (2000). *La distribution d'échantillonnage de l'estimateur du niveau d'habileté en testing adaptatif en fonction de deux règles d'arrêt : selon l'erreur type et selon le nombre d'items administrés*. Thèse de doctorat inédite, Université de Montréal.

Raîche, G. (2004). Le testing adaptatif. Dans R. Bertrand et J.-G. Blais (Dirs), *Modèles de mesure : l'apport de la théorie des réponses aux ite*ms. Montréal : Presses de l'université du Québec.

Scallon, G. (2004) *L'évaluation des apprentissages dans une approche par compétences.* Montréal : Éditions du Renouveau Pédagogique.

Sodoke, K., Nkambou, R., Raîche, G., Riopel, M. et Lesage, M. (2007). Extending Moodle functionnalities to adaptive testing framework. *E-Learn 2007,* 476-482.

Stocking, M.L. (1996). *Revising Answers to Items in Computerized Adaptive Tests : A Comparison of Three Models* (Rapport N°: TES-RR-96-12). Princeton, N.J: Educational Testing Service.

Tardif, J. (2006). *L'évaluation des compétences. Documenter le parcours de développement*. Montréal : Chenelière Éducation.

Tousignant, R. (1982). *Les principes de la mesure et de l'évaluation des apprentissages.* Saint-Jean-sur-Richelieu : Éditions Préfontaine Inc.

Wainer, H., Dorans, N.J., Flaugher, R., Green, B.F., Mislevy, R.J., Steinberg, L. et Thissen, D. (2000). *Computerized adaptive testing: A primer. 2nd edition*. Hillsdale, N.J.: Erlbaum.

Weiss, D. J. (1983). Introduction. Dans D. J. Weiss (Éd.) : N*ew horizons in testing - Latent trait test theory and computerized adaptive testing* (p. 1-7). New York : Academic Press.

Weiss, D. J. et Kingsbury, G.G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurment*. *21*(4), 361-375.

Weld, D. S., Anderson, C. R., Domingos, P., Etzioni, O., Gajos, K., Lau, T. et Wolfman, S. (2004). Automatically personalizing user interfaces*. International joint conference on artificial intelligence (IJCAI) 2003*, 1613–1619, Acapulco, Mexico.

Wiley, D. A. (2000). Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. Dans D.A. Wiley (Dir). *The Instructional Use of Learning Objects* [En ligne]. Accès: http://reusability.org/read/

XML (2007). *Extensible Markup Language.* [En ligne] Accès: http://www.w3.org/XML/

XUL (2008). *XML User interface Language (XUL).* [En ligne] Accès: http://www.mozilla.org/projects/xul/

Bibliography

Carr, B. et Goldstein, I. (1977). *Overlays: A theory of modelling for computer aided instruction* (Technical Report AI Memo 406). Cambridge, MA: Massachusetts Institute of Technology.

Chang, W., Hsu, H., Smith, T.K. et Wang, C. (2004). Enhancing SCORM metadata for assessment authoring in e-learning. *Journal of Computer Assisted Learning*, *20*(4), 305-316.

Cheniti-Belcadhi, L., Braham, R., Henze, N. et Nejdl, W. (2004). A Generic Framework for Assessment in Adaptive Educational Hypermedia. *Proceedings of IADIS International Conference WWW/Internet 2004 (ICWI 2004)*, 397-404.

Commission scolaire de Laval. (2007). *Situations-problèmes* [En ligne]. Accès : http://www.cslaval.qc.ca/tic/francais/grel/sitprobl.htm

Costello, D. (2002). Overview of AICC and SCORM Standards. *NASA E-Learning* [En ligne]. Accès : http://nasapeople.nasa.gov/Training/elearning/AICC-SCORM_standards.pdf

Doyon, C. et Juneau, R. (1991). *Faire participer l'élève à l'évaluation de ses apprentissages*. Montréal : Éditions Beauchemin ltée.

Eggen, P. et Kauchak, D. (2004). *Educational psychology: windows on classrooms 6th ed*. New-Jersey: Pearson Education Inc.

GTN-Quebec (2008). *XML Normetic.* [En ligne] Accès: http://www.normetic.org

IMS Global Learning Consortium, Inc (2007). *QTI standard* [En ligne]. Accès: http://www.imsglobal.org/question/

Lesage, M. (2007). *Moteur d'administration en ligne des apprentissages basé sur des normes formelles d'objets d'évaluation*. Communication présentée au 75è Congrès de l'ACFAS, Trois-Rivières, Québec.

Meirieu, P. (1987). *Apprendre… oui, mais comment?* Thiron: ESF Éditeur.

Ministère de l'Éducation, du loisir et du sport, Gouvernement du Québec. (2005). *Cadre de référence en évaluation des apprentissages au secondaire, version de consultation.* Québec, Ministère de l'Éducation, du loisir et du sport, Direction générale de la formation des jeunes.

Sclater N. et Low, B. (2002). IMS Question and test interoperability : An idiot's guide. *CETIS Assessment Special Interest Group* [En ligne]. *Version 0.5*. Accès: http://www.scrolla.ac.uk/resources/s2/idiots_guide.pdf

Silva, L., Mustaro, P. N., Stringhini, D. et Silveria I. F. (2006). Using Conceptual Lattices to Represent Fine Granular Learning Objects through SCORM Meta-Objects. *The Electronic Journal of e-Learning, 4* (2), 141 – 148.

Sodoke, K. (2007). *Les normes d'apprentissage à distance*. Document inédit, Département d'éducation, Université du Québec à Montréal.

Thissen, D. et Mislevy, R.J. (2000). Testing algorithms. Dans H. Wainer, N. Dorans, D. Eignor, R. Flaugher, B. Green, R. Mislevy, L. Steinberg et D. Thissen (Dirs.), *Computerized adaptive testing: A primer (Second Edition)*. Hillsdale, NJ: Lawrence Erlbaum Associates, 101-133.

Varlamis, I. et Apostolakis, I. (2006). The Present and Future of Standards for E-Learning Technologies. *Interdisciplinary Journal of Knowledge and Learning Objects, Vol.2*, p. 59-76.

Website

SOAP. *Simple Object Access Protocol (SOAP)* [En ligne]. Accès: http://www.w3.org/TR/soap/