# Creating personalized assessments based on learner knowledge and objectives in a hypermedia Web testing application

**Fotis Lazarinis, Steve Green, Elaine Pearson**

School of Computing

University of Teesside

Middlesbrough Tees Valley, TS1 3BA, UK

f.lazarinis@scm.tees.ac.uk, s.j.green@tees.ac.uk, e.pearson@tees.ac.uk

**Abstract**

This paper presents an adaptive Web testing system where the adaptation of the testing procedure relies on the performance, the prior knowledge and the goals and preferences of the test participants. The main aim of the paper is to discuss the utility of the adaptive assessment tool. Therefore, the paper presents the authoring process of a number of adaptive assessments, where the educators utilized different adaptive rules to create summative and formative assessments. The paper presents the employed adaptive rules and summarizes the key points of the pilot study.

**Keywords:** Adaptive Testing, Personalization, Multimedia Assessments, Assessment Information System, Adaptive Hypermedia, Evaluation

## 1. Introduction

Assessment plays an important role in every learning and teaching activity. What is assessed and how is assessed influences the instructional process in several ways. Computer based assessment tools are information systems that automate the testing process, making it also more interactive and enriched with various media types. Further, they attempt to overcome the barriers of time and distance by being available through the Internet at any time. Several fixed-length and adaptive computerized tests have been developed. Computerized adaptive testing (CAT) systems are computerized tests, aiming to accurately reflect the instructional level of each student with a shorter number of queries tailored to the characteristics of the individual learner (Wainer, 2000; van der Linden & Glas, 2000). The existing computerized adaptive testing (CAT) systems base their adaptation mainly on the learner's performance using statistical models, which can be considered as restrictive from a pedagogical viewpoint. As it will be discussed later, some novel adaptation approaches for testing, which seem more pedagogically promising, have already been suggested (e.g. Brusilovsky & Sosnovsky, 2005; Sitthisak et al., 2008).

In this paper, we first present a Web testing system where the adaptation is based on multiple criteria, such as the performance, the goals and the previous knowledge of the test participants. Each criterion is customized by the test authors and multiple rules can be applied at certain points of the test. These rules adjust, eventually, the testing procedure to each learner and could be used in order to make useful deductions at the end of the test. The emphasis of the current paper is on presenting the authoring process and some real testing cases where the participated educators authored different adaptive assessments in different testing settings.

The remaining of the paper is organized as follows. Section 2 analyzes a number of studies related to adaptive e-assessment tools. Section 3 presents the domain, user, and adaptation models of the adaptive testing tool and presents a prototype of a tool. Section 4, presents ten

authentic adaptive assessments and analyzes the adaptive rules with respect to the aims of the educators. The opinions of the educators who created the adaptive tests are also presented. Section 5 summarizes the work and gives directions for further research.

## 2. Background to adaptive testing tools

At the beginning of the 1970s, some works pointed out that the creation of more flexible tests could solve some of the problems of fixed length tests. In Lord (1970), the theoretical structure of an adaptive test was defined which could be administered to groups but tailored to each individual. Trying out adaptive tests was not possible until the beginning of the 1980s, when computers became more powerful and less expensive. The philosophy underlying an adaptive test is described in Wainer and Mislevy (1990). The authors mention that adaptive tests attempt to reproduce the practices of examiners who adjust the sequence of their questions based on the responses of the students.

A Computer Adaptive Test (CAT) can be defined as a test administered by a computer where the presentation of each item and the decision to finish the test are dynamically adapted based on the answers of the examinees. CAT tools aim to accurately reflect the instructional level of each student and they are used mainly as skill meters presenting the overall learner's score on a subject and a pass/fail indication. More specifically, test items dynamically adjust to a student's performance level, and as a result, tests are shorter and test scores tend to be more accurate (Thissen and Mislevy, 2000).

A number of computerized adaptive testing tools have been implemented by academic institutions, e.g. SIETTE (Conejo et al., 2004), and international companies for specific examinations (Microsoft CAT, 1999; GRE Exam, 2006), and have been extensively utilized in recent years.

In the Lisp Tutor ELM-ART (Weber and Specht, 1997; Weber and Brusilovsky, 2001) there is a testing component that uses information about the performance of the student in previous tests to select the next question to ask. Medtec (Eliot et al., 1997) is a web-based intelligent tutor for basic anatomy. Tests are generated automatically based on an overlay student model.

In (Lee and Wang, 1997) a hypermedia learning system is presented. In this tool, a fuzzy approach is used to diagnose the student's answers and to create and update an overlay student model. This student model is used to select the problem to be presented to the student.

The selection of problems in all adaptive testing systems is based on the same principle: select problems according to their difficulty and the student's performance; the better the performance, the more the difficult question is presented. The main advantage of a CAT over a traditional computerized test is efficiency. The CAT systems can determine a person's score with fewer questions, sometimes reducing the length of the test by 60% or more. Nevertheless some limitations of CAT systems have been identified in the literature. First they base their adaptation solely on the performance which limits their use for alternative pedagogic purposes (Wainer, 2000; Wise and Kingsbury, 2000). It has also been argued that it is impossible to feed an operational adaptive test with brand-new, unseen items; all items must be pre-tested with a large enough sample to obtain stable item statistics (Wainer, & Mislevy, 1990). As it has been observed, a few items are overexposed while several questions are under-selected, so test creators argue that constantly enriching the question database is of no practical use since the majority of questions are not utilized (Leung et al., 2003). Further, in some simulated evaluations it was shown that students with an average knowledge on the topics of the test may not be appropriately assessed in CAT systems (Abdullah Chua & Cooley, 2002). In other words, the system cannot estimate accurately the true knowledge of students with average knowledge. Some novel approaches have been proposed to alleviate the computational complexity and resolve the problem of item exposure in CAT tools (Huang et al., 2009).

## 2.1 Different approaches in adaptive testing tools

Different types of adaptivity in assessment systems can be found in QuizPACK (Brusilovsky & Sosnovsky, 2005) and QuizGuide (Sosnovsky, 2004). These tools support self-assessment of programming knowledge with the aid of Web-based individualized dynamic parameterized quizzes and adaptive annotation support (Brusilovsky, 2001). Multiple versions of the same queries are offered to learners. Also they have the opportunity to see the right answers and try again the same question with different parameters. Adaptive annotation means that various visual cues are attached to the links on the current page pointing in quizzes in order to help users select and execute the most relevant quiz. The tools described in these papers are domain dependent and are basically possible in domains such as mathematics, physics and programming.

Another adaptation technique, adaptive questionnaires (Kehoe & Pitkow, 1996), has been used mainly in computer-assisted Web surveys. This method causes the generation of a dynamic sequence of questions depending on learner's responses reducing the number and complexity of questions presented to users. The techniques used in adaptive questionnaires are interesting because the adaptation is decided by the designer of the survey and therefore multiple factors could be integrated into the questionnaires.

A set of competencies are defined in (Sitthisak et al., 2007) and the next assessment stages rely on the competencies that an individual possesses. The set of competencies are defined by a number of parameterized attributes and thus the competencies can be modified for different domains. The paper actually presents only a framework for defining adaptive assessments and not a tool. It further refers to some disadvantages of CAT systems to estimate accurately the knowledge of learners. This work is extended in (Sitthisak et al., 2008) where the authors present a prototype of a tool for automatically creating a number of questions for a required competency based on the associations of the questions to various competencies.

An adaptive assessment framework as part of the diagnostic module in the Adaptive Educational Hypermedia System INSPIRE (Intelligent System for Personalized Instruction in a Remote Environment) is presented in (Gouli et al., 2002). The main goal of the assessment procedure is to assess the learner's knowledge and to give learners a chance to keep track of their learning progress. A pre-test is used as the initial estimation stage where questions may be associated with one or more concepts. A number of parameters are associated with each question such as the level of difficulty, the level of performance, the number of times that the question has been answered correctly or incorrectly by any learner etc. Based on these data the next questions are decided.

The early work on a rule-based testing system is discussed in (Tzanavari et al., 2004). The system presented refers to an architecture for creating rule-based adaptive assessments with limited customization options. The ideas presented are interesting but the papers focus on the learner modelling capabilities and the rules which concern the performance of the learner only, as in the CAT systems. In (León et al., 2008) the learning content is re-adapted based on the results of an initial evaluation after the presentation to students. The results of the initial evaluation are also used to decide the entry difficulty level for the subsequent evaluations.

The main conclusion of the literature review is that researchers explore new strategies for personalization in assessment tools. The proposed techniques aim at increasing the pedagogical value of the adaptive testing applications and at providing more efficient ways for individualizing the assessments to the instructional aims of the teachers and the goals of the learners.

Towards these research directions, we present an adaptive testing system where the adaptation of the testing procedure relies on the performance, the prior knowledge and the goals and the preferences of the test participants. Test authors are able to customize different rules to meet their diverse instructional aims. The system attempts to reflect the beliefs of the individual teacher who understands better his/her students and eventually can adapt the process to their

specific attributes or difficulties. Tests are coded using IMS QTI (2006) to support the interoperability of the testing items.

## 3. An adaptive Web testing tool with customizable rules

In this paper a framework (see Figure 1) for creating adaptive tests is presented, where the adaptation of the assessments relies on factors such as the knowledge, educational background, goals, preferences and performance of the learners. This conceptual structure considers a number of user-customizable rules where the conditions and the actions can be parameterized by educators to create personalized assessments. The Web tool implementing the framework supports the authoring of adaptive and non-adaptive tests, through the customization of pre-defined rules. The authors can enter the testing data, which are divided into sections and to apply conditional rules at specific items for forcing advancement to next sections or re-attempt the wrongly answered questions. The tool also supports the population of learner profiles with their identification details, their initial knowledge in the tested topics, their aims and their educational background. These learner data influence the application of adaptive rules, in addition to the performance of the learner in a specific test.
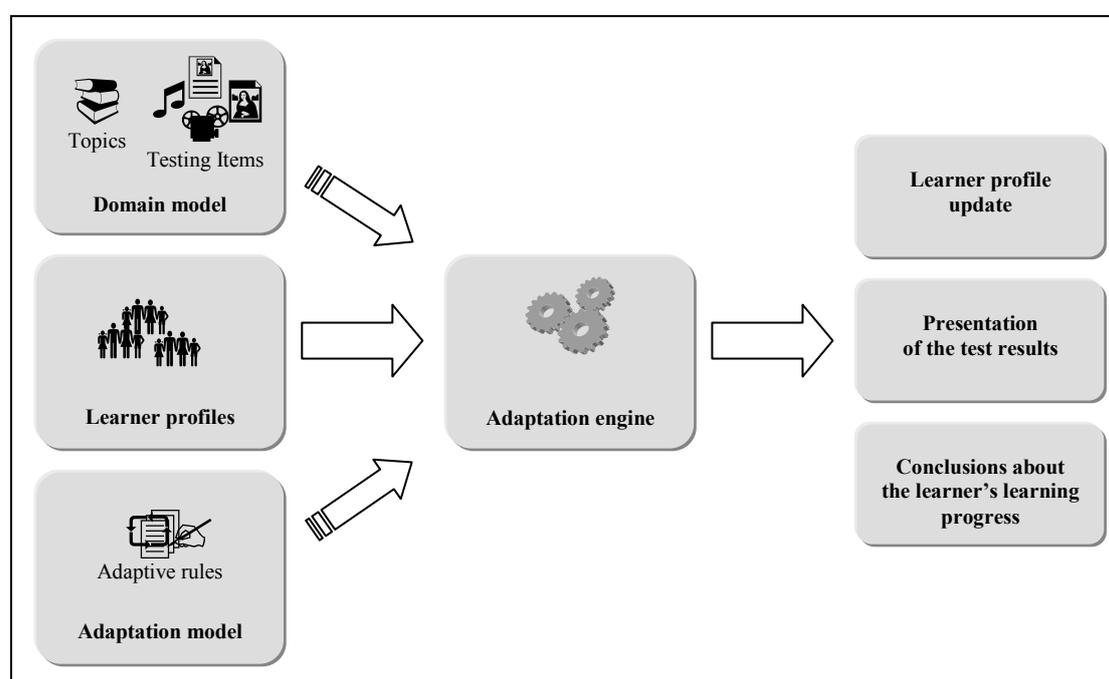


**Figure 1.** Main components of the adaptive testing tool

Typically, adaptive e-learning tools consist of four parts that work closely together (De Bra et al, 1999). The *domain model* maintains the topics, concepts and other fragments that are used in forming the lessons. *User models* contain information about the learners which vary from demographics (name, address, etc) to their current knowledge and to their educational data. The *adaptation model* is a part in the adaptive hypermedia tool model defined as a collection of rules that define how the adaptation must be performed. The rules are used for updating the user model through the generated relationships between the concept and the existing learner knowledge. The final part, the *adaptive engine*, performs the actual adaptation. The adaptation model describes the conditions and the actions on which the presentation of the information is based and the adaptive engine implements these rules.

In the presented adaptive system the domain model consists of the topics and the testing items which are adaptively presented to the learners. The user model component of the exemplar adaptive learning tool corresponds to the learner profile module of our adaptive testing tool. The adaptation model consists of a set of customizable if-then-else rules concerning the performance, the knowledge and the goals of the test participants. The adaptive engine takes as input the testing items and their associated topics, the learner characteristics and the

employed rules and presents the tests in an adaptive mode to the participants. It also updates the learner profiles with the current knowledge estimations and presents statistics about the tests taken. It also attempts to make a number of useful assumptions about the true knowledge and the potential learning difficulties of the learners. These deductions are useful for both educators and learners for helping them appreciate their true knowledge.

### 3.1 Domain model

*Topics*

The topic hierarchies in the presented adaptive testing system are sets of associated concepts. Concepts in the domain model may be atomic or composite containing other atomic or composite concepts. A composite concept is made of one or more sub-concepts. The nodes of the hierarchy without sub-concepts are atomic concepts. These hierarchies or graphs are used by the system to associate testing data to specific concepts and then they are overlaid into the learner model. This means that each testing activity is associated with a concept and the estimated knowledge level related to this concept is recorded in the learner model. Topic hierarchies in our system are coded using Topic Maps (Topic Maps, 2002), which is an ISO standard for the representation and interchange of knowledge. A topic map can represent information using topics, associations, and occurrences.

*Testing data*

Tests are divided into sections consisting of questions of various types (true/false, single/multiple choice, order, etc). Questions and tests are online and they reside on the server's side (see Figure 2). Tests and questions are coded using the IMS QTI XML e-learning standard (IMS QTI, 2006) for supporting re-usability of the tests in other learning tools. Each section can be associated with one or more topics and one or more sub-topics. From a pedagogical point of view this is essential, as it will allow the association of student knowledge to specific concepts. Further, as seen in Figure 2, educators are able to define additional assessment and section options such as the duration of the test, the navigation mode, comments to candidates, feedback, etc. These metadata correspond to specific elements of the QTI XML standard and are important in further adapting the assessment to the aims of the educators. The author may, further, associate multimedia content to each query, such as images, video or sounds. These files are referenced in the query's XML file and are stored in a separate directory so, that they can be shared across different queries.

**Figure 2.** Assessment and question editing

### 3.2 Learner profiles

Learner profiles consist of a set of attributes which describe the personal data of learners, their formal education and previous knowledge, their goals and their evaluation data. In the adaptive testing tool presented in this paper, learner profiles must maintain information about the knowledge of the learners on specific topics and data about their testing activities. These data will be used to adapt the testing activity and could be also presented at the end of a testing activity in order to provide detailed descriptions of the progress of the student.



6

**Figure 3.** Learner profile management

Learner profiles are managed with the aid of a Web tool (see Figure 3). In general, the tool supports the following set of actions:

- Create new learner profiles.

- Update learner profiles.

- Present parts of the profiles.

Educators can create new learner profiles which can be altered at any point by logging into the tool and changing the respective parts. Presentation of the learner profiles can be refined according to specific criteria, such as the last educational activity, or previous formal education, or the goals and preferences of the student.

The data are divided into five major categories. The first category, i.e. "demographics", contains the identification data of the learners. The purpose of the "Education and previous knowledge" category is to capture the knowledge of the learner on specific topics through their formal or continuing education. Therefore the details of their secondary education and possible university or informal educational details are included.

In adaptive education hypermedia systems, topics are overlaid in the learner profile (De Bra et al., 2004). That is, the knowledge of the learner is associated and measured with respect to specific topics of the topic hierarchy. Therefore, the knowledge of the learners to topics should be included in the learner profile. The knowledge is noted as *none*, *low*, *good*, or *very good*.

The goals of learners are limited at the moment to whether they wish to reach specific scores on assessments or whether they wish to try specific questions which either match their knowledge level or are of specific difficulty. Learners can associate one of these goals per assessment or they can rank their goals from the most important to the less significant. Then, during the assessment, the goals that do not collide will be applied starting from the most significant.

The preferences concern the presentation of results at the end of the assessment or the presentation of the test in limited resource environments. In general, the goals and preferences in the prototype of the learner management tool are only indicative of the potentials of the adaptive application. The purpose of these categories of data is to demonstrate that it is

possible to include specific learner characteristics in personalized tests and to test whether test participants and educators want this type of adaptivity. Different goals and preferences could be added in future versions of the application.

### 3.3 Adaptation model

During the construction of a test, the educators are able to customize one or more rules and apply them to specific points in the test (see Figure 4). Each adaptation rule has an *initiation point*, a *condition* and a specified *action*. Initiation points are set on specific questions of the assessment and testing routes change based on whether the conditions are met.



**Figure 4.** Application of adaptive rules

**Initiation and conditions**

Tests consist of sections and questions. So, the initiation point could be at the start of the test or at the start of a section or at a specific question of a section.

The supported conditions concern:

- the performance of the learner,

- the prior knowledge and formal education of the learner,

- the goals or the preferences of the learner.

*Performance*

The performance of the learner can be decided, at any point of the test, based on the percentage of the correctly answered questions. The scoring procedure takes into account the difficulty of the questions or the number of attempts per question in order to adjust the mark for each question. The more difficult a question the higher mark a learner will take. The score for each question is defined by the educator during the test construction phase. Failed attempts for a question may lead to a deduction in the score, e.g. one mark for each failed attempt. Multiple attempts and penalization for failed attempts are also up to the educators to decide during the authoring of the adaptive tests.

*Prior knowledge and formal education*

The prior knowledge of a learner on specific topics can be indicated as *none*, *low*, *good*, or *very good* during the creation of learner profiles. Formal education can be used for initializing automatically the knowledge level of a learner to the topics related to their studies. Later, the knowledge level on the topics, negotiated on the tests taken by learners, is updated according to their performance based on specific limits set in the assessment.

*Goals and preferences*

The goals originate from the purpose of the learners taking a test and other factors such as their current knowledge. When preparing, for example, for a final exam, then some students might want to try all the available questions while some others with a higher knowledge might want to try only the difficult questions. Therefore the system needs to take different aspects into account in order to accommodate the diverse goals of learners.

The preferences of the test participants may relate to the execution environment or potential disabilities that might have or on how they wish the results to be presented. So preferences like specific colour combinations, font sizes, or to be shown with analytical results at the end of a test could be included in the system to reflect the needs of users.

**Actions**

The previous section described the supported conditions and explained the factors on which they are based. This section describes what actions follow the satisfaction of a condition. These actions allow learners to move on to the next section or question or to re-try specific parts of the assessment. The supported set of actions is:

- start assessment on question

- start assessment on section

- move to question

- move to section

- re-try specific questions

- re-try specific section

- re-try assessment

- show questions of specific difficulty level

- show questions that match the knowledge level

This set of actions allows users with a high knowledge level or with a high performance to proceed faster and to give a second chance to learners with lower performance. Also it takes into account the goals of the learners by asking them to decide on the difficulty level of the questions.

The adaptation management tool (see Figure 5), allows educators to visually customize a number of predefined rules and apply them to certain questions. Educators can select a test and then apply some adaptive rules on specific points. The tool loads the test structure and visually displays it allowing the educators to customize the available criteria.

Educators are able to apply different rules at different points. For instance, in one section they could base their adaptation on performance and in another section of a test could be customized based on another factor. Or the same rule could be adapted in a different mode. For example, Figure 5 depicts the application of the rule "In each section, if the test participant has good or higher knowledge on the tested topic, then show questions of medium or high difficulty".

### 3.4 Adaptive engine

The run time environment, which can also be used as a preview tool during the authoring phase, takes as input the assessment data and presents each query to the user. The test participant is informed about the topic of the question and is presented with the associated feedback at the end of a question. The results of the question may be presented at the end of the question or at the end of the assessment, depending on the restrictions set during the authoring of the assessment. At the end of the assessment procedure, learners are presented with analytical statistics and inferences about their progress. These data are stored for later reference by learners and educators. Educators may have named access to all the data of their learners and inferences automatically made by the system will be presented to them.

### 3.5 Synopsis

The previous sections described an adaptive Web testing tool where the adaptation criteria can be authored to accommodate different instructional aims. The proposed adaptive system differs from the previous statistics-based adaptive testing systems as it takes into account the goals and the prior knowledge level of the test participants, in addition to their performance. So it is possible to apply adaptive rules like "show only questions that match the knowledge level of the test participant" or "re-attempt the questions which were wrongly answered".

The obvious advantage of this approach to personalization of the tests is the flexibility offered to both learners and educators. Educators are able to reflect their personal instructional experiences and aims, in order to create tests more tailored to the characteristics of specific learners or to specific groups with special characteristics. Learners can take advantage of their previous knowledge or their current goals and be examined in shorter tests with more focused items. This is especially useful in formative assessments. Formative assessment is often performed at the beginning or during a training program, thus providing the opportunity for immediate evidence for student learning in a particular course or at a particular point in a program. Quick estimation of a learner's knowledge and competence is therefore possible.

In statistics-based adaptive testing tools formal assessments are based on mathematical formulas which differentiate the presented items according to the answers of the students. These tests, however, do not take into account specific aims of the educators. In our adaptive framework, formal assessments can also rely on the performance only, but the actual parameterization is left to the test designers who can reflect their particular instructional aims. For example, different performance thresholds can be set for specific learner groups or for tests of increased difficulty. These thresholds will be identical for all students making the

testing process common to all participants. The integration of the beliefs of the test-developers into the system, even in formal assessments, is another advantage of our system with comparison to computerized adaptive tests.

## 4. Authoring of adaptive tests

The previous section presented an adaptive testing tool and focused on the authoring capabilities of the system. It presented the components of the application and discussed the customisation of the adaptive rules. In this section we analyse the experiences of ten educators who used the system to create personalized assessments for different topics. The current evaluation focuses on presenting the applied adaptive rules and the opinions of the educators about the system. The basic aims of this evaluation were to record the opinions of educators of various disciplines towards a rule-based adaptive tool and to evaluate the utility of the existing rules and the needed extensions.

The ten educators participated in the current evaluation were high school teachers of various specialties (2 Computer science teachers, 2 Teachers of history, 1 English language teacher, 2 Teachers of Mathematics, 1 Teacher of Physics, 1 Teacher of Geography, 1 Teacher of Greek linguistics) and keen on enhancing their teaching with the use of technology. They worked as teachers for a period of 3 to 11 years with an average of approximately 6.5 years. They participated in the evaluation experiments voluntarily and all of the participants had at least basic computer handling abilities. Three of them had used non-adaptive e-testing tools like HotPotatoes (http://hotpot.uvic.ca) in the past. The functionality and the adaptive capabilities of the system were explained to each educator for approximately 30 minutes. Then the participants were asked to create a realistic assessment for their high school classes.

Table 1 shows the number of sections that were created per assessment and the topic of each assessment. As seen the educators created different number of sections and questions per section. The majority of the 329 questions were true/false, single and multiple choice and fill-in-the gaps. The average completion time for each test was 70 minutes. As it was noted the majority of the teachers defined only the score and the difficulty of each question. The other options (e.g. attempts) were left with their pre-defined values. In 29 questions the educators provided some feedback, which was either a Web location for further material or short explanatory texts. A simple topic hierarchy composing of composite topics was used for this test, e.g. Computer science/Introduction to algorithms, History/Modern history, etc.

After the construction of each test, a questionnaire was filled out by each educator and a short interview was performed with each individual participant. In the next paragraphs each assessment is described in more detail and the employed adaptive rules are analyzed.

**Table 1.** Topics and sections per assessment

| Educator # | Topic | No of questions per section | | | |
|---|---|---|---|---|---|
| | | Sec. #1 | Sec. #2 | Sec. #3 | Sec. #4 |
| 1 | Computer Science, Introductory algorithmic concepts | 15 | 15 | 15 | - |
| 2 | History, Greek modern history | 10 | 10 | 10 | - |
| 3 | History, Second World war | 12 | 12 | 12 | 12 |
| 4 | English language, Use of tenses | 8 | 8 | 8 | - |
| 5 | Mathematics, Calculations (additions, subtractions, multiplications, divisions, percentages) | 8 | 8 | 8 | 8 |

| 6 | Computer Science, Conversions between arithmetic systems (binary, decimal, octal, Hexadecimal) | 6 | 6 | 6 | 6 |
|---|---|---|---|---|---|
| 7 | Physics, Electromagnetism | 8 | 8 | 8 | - |
| 8 | Greek linguistics, Grammar rules | 6 | 6 | 6 | |
| 9 | Geography, Environmental systems | 9 | 9 | 9 | 9 |
| 10 | Mathematics, Statistics | 10 | 10 | 8 | - |

## 4.1 Rules used in assessments

*Assessment #1*

The first assessment consisted of three sections of 15 questions each. In each section the first five questions were easy, the next five items were of medium difficulty and the last ones were difficult. All the questions were either true/false or single choice. The educator applied the following two rules:

Rule 1: If the knowledge level of the test participants is good or very good then start the assessment on the first medium difficulty question.

Rule 2: At the end of each section if the score is greater or equal to 80% then move on to the first medium difficulty question of the next section, otherwise continue on the first question of the next section.

Based on the short personal interview, performed on the completion of the test, the goal behind there rules was to help student with a higher knowledge to proceed faster. The first rule was based on the prior knowledge on the topics of the test, as this is stated in the learner's profile. The second rule is based on the current knowledge estimation, as this is decided by the achieved score in each section.

*Assessment #2*

The second assessment, related to Greek modern history consisted of three sections of ten questions each. In this case, the first section contained the easy questions, the second part of the test contained the medium difficulty questions, and the last section the difficult questions. The employed rules are:

Rule 1: Ask the learner to decide whether s/he wants to start on the easy or medium difficulty questions.

Rule 2: Re-try the erroneous questions of each section if the score is lower than 80%.

As it was explained to us, this test was a formative assessment aimed to help students prepare for their final trimester examination. Therefore students were let to decide on which section to start their test but in case they had a lot of wrongly answered questions, they were given the opportunity to re-try the erroneous items in order to understand their misconceptions.

*Assessment #3*

The next assessment concerned again historical issues and consisted of four sections of 12 questions each. The first section consisted of easy questions, the next two sections of medium difficulty questions and the last one contained difficult questions. The third educator adjusted

the test primarily on the achieved score in each section in order to allow students who were well prepared to proceed faster. The employed rule was:

Rule 1: On the eighth question of each section, if the score is 100% then move on the first question of the next section.

*Assessment #4*

The fourth educator constructed three sections of eight questions each. Questions did not follow any specific pattern, that is in each section the easy, medium and difficult questions were mixed up in an undefined form. The only rule applied was:

Rule 1: Re-try the wrongly answered questions of each section.

The educator explained to us that each question is important as it deals with a different grammatical phenomenon and that it was important that all questions had to be attempted by all students. Also, in case of erroneous answers the teacher argued that the students have to re-try the questions in order to understand their mistakes. This was primarily a formative assessment.

*Assessment #5*

In this test, the mathematician constructed four sections. The first one contained easy additions and subtractions, the second contained additions and subtractions of negative and decimal numbers and the third and fourth sections contained multiplications, divisions and percentage calculation. The questions of the second section were defined as medium difficulty questions and the questions of the last two sections as difficult questions. The employed rules were:

Rule 1: On the sixth question of the first section, if the score is 100% then move on the first question of the next section.

Rule 2: On the sixth question of the second section, if the score is 100% then move on the first question of the next section.

Rule 3: Re-try the wrongly answered questions of the third section.

Rule 4: Re-try the wrongly answered questions of the fourth section.

The educator wanted to help students to proceed faster in easy or medium difficulty additions and subtractions and to try all the multiplications, divisions and percentage calculations. That is why the third and fourth rules were employed.

*Assessment #6*

The sixth assessment consisted of four sections of six medium difficulty questions each. Each section concerned a different arithmetic system. The employed rule concerned primarily the performance:

Rule 1: On the fourth question of each section, if the score is 100% then move on the first question of the next section.

The purpose of the educator was to help students with a high knowledge to proceed faster but also to try questions from all arithmetic systems.

*Assessment #7*

In this case, the educator created three sections with four medium difficulty and four difficult questions each. The applied rules were:

Rule 1: If the knowledge level of the test participants is very good then start the assessment on the first difficult question of section one.

Rule 2: At the end of each section, if the learner has replied correctly all the difficult questions then move on the first difficult question of the next section.

The rationale behind this assessment was to help students with a high knowledge to complete their assessment faster.

*Assessment #8*

The teacher of Greek linguistics constructed three sections of six questions each. No adaptive rule was applied in this case, as the educator argued that this was a summative assessment and all students had to reply the same number of questions.

*Assessment #9*

The four sections of this assessment consisted of nine questions, three of each difficulty level. The applied rules were:

Rule 1: Ask the learner to decide the difficulty level of the questions.

Rule 2: Show questions that match the knowledge level of the learners.

These two rules were mutually exclusive, meaning that if the learners want to try specific questions to meet their goals, then the second rule will be applied. Otherwise the system will show only the questions that match their knowledge level.

*Assessment #10*

The last assessment concerned mathematic concepts related to the calculation of percentages or other similar exercises. The first section had easy questions, the second had medium difficulty questions and the last one difficult questions. The applied rule concerned primarily the performance of the students:

Rule 1: On the sixth question of each section, if the score is 100% then move on the first question of the next section.

The mathematician argued that this is a summative assessment and only students with a good knowledge, i.e. a high performance in the current test setting should proceed faster.

### 4.2 Questions to the educators

Following the construction of the assessments and the discussion with each educator, a short questionnaire was distributed to the participants in order to record their opinions about the system. The questionnaire consisted of ten questions scored in a five-point Likert scale (i.e. strongly disagree to strongly agree). After the completion of the questionnaire a discussion with all the participants was carried out to comment on their replies. The following paragraphs present the questions and comment on the replies of the participants.

*The tool was easy to use*
All the participants strongly agreed (100%) that the tool is easy to use mainly because of its

simple Web interface, through which users can edit their questions and apply the desired rules. Also the fact that all the interfaces follow the same convention and aesthetics helped them in mastering the interface very quickly.

*The system is useful*

We asked the participants to comment on the significance of the system. All the participants strongly agreed (100%) that the system is indeed useful as it allows them to assess the knowledge of the students by applying different testing strategies. This is one of the most important questions of the evaluation as it helps us understand the importance of the system. Although, the number of the participants is not as large as we would wish the agreement of the educators is an indication of the system's significance.

*What is the most important feature of the tool?*

The next question was to identify the most important feature of the adaptive testing tool. We provided the users with the following list of choices:

1. Application of different adaptive rules
2. Easy to learn and to use
3. Supports different question types
4. Attractive interface
5. Ability to create shareable learning content
6. Other (please specify): ………………….

These choices were shown in a different order to each user to avoid the order effect bias. Nine of the ten (90%) educators indicated the ability to apply different rules is the most important feature of the tool, as it allows them to support different testing strategies. One of the educators selected the "easy to learn and to use" as the most important feature of the tool.

*Authoring of adaptive rules was easy*

All the participants strongly agreed (100%) that the interface for editing adaptive rules was easy to use and the options were self-explanatory.

*I could edit all the question types*

Four of the ten participants strongly disagreed (40%) on this statement. The other six (60%) were either neutral or agreed with the question. The participants who disagreed noted that they cannot enter special characters or other symbols needed for describing mathematical symbols, for example.

*Most of the attributes were useful*

Although most of the participants left several of the assessment's attributes undefined during the assessment construction, they all considered (100% - agreed or strongly agreed) the existed attributes as useful. They argued that the existence of several attributes could allow them to better adjust the assessment to the needs of their students and to their instructional aims.

*The technical details were hidden*

All the tutors strongly agreed (100%) that the technical details of the system were hidden. The instructors argued that the underlying technologies should unquestionably be transparent to the educators, in order for e-learning tools to be adopted by educational staff with minimum technical knowledge.

*The adaptive rule set needs expansion*

As expected, all the educators strongly agreed (100%) that the adaptive rules need some extensions to cover other aspects of the learners, such as their cognitive ability, possible disabilities, learning style, computer environment, etc. Also they would like the existing rules to be more flexible. However, they also pointed out that a tool which might be too complex, may be unusable in the long run as the educators may not be willing to spend too much time in mastering it. The replies and the comments of the educators to this question will be addressed in our future research plans.

*The system needs more features*

Most of the educators argued (20% neutral, 40% agreed, 40% strongly agreed), that the tool needs more features. They suggested some new features such as alternative query text, links to external resources, different marking schemes, help on demand during authoring, etc.

*I would use the system in my classes*

We asked the users if they are willing to continue using this tool in their classes. Although they were some worries about the stability of the system, since it is the result of ongoing research, all the educators (100%) were willing to use the tool in their next courses. The main reason is that they feel that with such a flexible tool they can diagnose the real knowledge of their students better, with a shorter set of questions.

*4.3 Synopsis*

In the previous paragraphs we described the authoring process of ten different adaptive assessments. No significant problems were observed during the evaluation experiments. In order for the educators to meet their diverse instructional aims, the authors of the assessments applied different adaptive rules, by customizing a set of pre-defined rules. In most of the cases the prior knowledge and the user goals influenced the testing procedure more than the performance of the testers. In the summative assessments, adaptation was based primarily on the current knowledge level estimation and on the performance of the learners. In the formative assessments, adaptation was based either on the previous knowledge of the learners or on their goals. The learners are also given the opportunity to re-attempt the wrongly answered questions in order to understand their errors and their possible misconceptions.

As it can be concluded from the diversity of the applied rules in the ten assessments, the presented adaptive application is quite flexible to accommodate different pedagogical aims and testing strategies. Through their replies, the test authors validated that the system is actually useful and easy to use and although it needs some extensions, it could be used in their future testing activities. According to the test participants, the main advantage of the application is the support of different testing strategies, which makes it appropriate for a number of different testing settings. The needed extensions concern the adaptation options, the authoring capabilities for special question elements, and the available question customization options.

Although the number of the participating educators is not as expanded as we would wish, the diversity of their specialties, the variety of their instructional aims and the variation in the application of adaptive rules make the findings, in general, solid. Nevertheless, an expansion of the user base is definitely needed to overcome any limitations caused by the small number of participants.

**5. Conclusions**

In this paper we present the modules and the authoring process of an adaptive hypermedia testing tool where the adaptation criteria can be authored to accommodate different

instructional aims. The aim of the present work was to investigate the usefulness of the system through its utilization by the educators. Authors of assessments were able to customize pre-defined rules which concern the performance, the goals and the previous knowledge of the test participants. The data of learners are stored in individual profiles so as to adapt the learning process according to their characteristics. The data maintained in the learner profiles vary from contact details to the knowledge in specific topics, to formal education and to the goals of the users. Based on these data, educators are able to apply different rules into the same assessment, which eventually create different assessment paths.

The proposed adaptive system differs from the previous computer adaptive systems as it takes into account the goals and the current knowledge level of the test participants, in addition to their performance. So it is possible to apply adaptive rules like "show only questions that match the knowledge level of the test participant" or "re-attempt the questions which were wrongly answered". The obvious advantage of this approach to personalization of tests, is the offered flexibility to both learners and educators. Educators are able to reflect their instructional experience in order to create tests more tailored to the characteristics of their learners and to motivate and help different student categories. Learners can take advantage of their previous knowledge or their current goals and be examined in shorter tests with more focused items.

Overall, the general feeling of the evaluation experiments was that educators yearn for more adaptable and testin testing environments. The evaluation showed that the system is useful and flexible. However, it could be further improved to provide more facilities to its prospective users. It could be improved by extending the adaptation model and by further automating the application of rules by using step-by-step wizards. This could help authors to faster and more accurately apply adaptive rules to their assessments. Another service which could improve the system is the integration of ready-to-use templates with pre-existing adaptive testing structures. These templates need to be designed with the assistance of experienced educators to cover various testing cases. Finally, the development of an efficient help system could further assist the educators to solve potential problems arising during the utilization of the tool.

Apart from the above possible expansions, in the future the tool will be used in assessing the knowledge of students in various domains in order to value the importance and the suitability of the tool for various assessment cases. Also the tool needs to be utilized by more educators to see how the tool supports and enhances their teaching and assessment demands and to identify potential improvements.

## References

Abdullah Chua, S., and Cooley, R. (2002). Using Simulated Students to Evaluate an Adaptive Testing System. *IEEE Proceedings of the International Conference on Computers in Education*, 614-618.

Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11, 87-110.

Brusilovsky, P., Sosnovsky, S. (2005). Individualized exercises for self-assessment of programming knowledge: An evaluation of QuizPACK, *Journal on Educational Resources in Computing (JERIC)*, 5(3), http://doi.acm.org/10.1145/1163405.1163411

Brusilovsky, P., Sosnovsky, S. (2005). Individualized exercises for self-assessment of programming knowledge: An evaluation of QuizPACK, *Journal on Educational Resources in Computing (JERIC)*, 5(3), http://doi.acm.org/10.1145/1163405.1163411

Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J., and Ríos, A. (2004). SIETTE: A Web-Based Tool for Adaptive Testing, *International Journal of Artificial Intelligence in Education*, 14(1), 29 - 61.

Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J., Ríos, A. (2004) SIETTE: A Web-Based Tool for Adaptive Testing, International Journal of Artificial Intelligence in Education, 14(1), pp. 29 - 61.

De Bra P., Aroyo L. and Cristea A. (2004). Adaptive Web-Based Educational Hypermedia, in Levene and M. Poulovassilis A. (Eds.): *Web Dynamics, Adaptive to Change in Content, Size, Topology and Use*, Springer, pp. 387-410.

Eliot, C., Neiman, D., and LaMar, M. (1997). Medtec: A Web-based Intelligent Tutor for Basic Anatomy. *Proceedings of WebNet'97, Second World Conference of the WWW, Internet and Intranet* (pp. 161- 165). ACCE.

Gouli, E., Papanikolaou, K., Grigoriadou, M. (2002) Personalizing Assessment in Adaptive Educational Hypermedia Systems, Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, LNCS 2347, pp. 153 - 163.

GRE Exam, (2006) *Kaplan GRE Exam*, 2007 Edition: Premier Program.

IMS QTI (2006). Question and Test Interoperability, Available at: http://www.imsglobal.org/question

IMS QTI (2006). Question and Test Interoperability, Available at: http://www.imsglobal.org/question

Kehoe, C., & Pitkow, J. (1996). Surveying the territory: GVU's five WWW user surveys. *The World Wide Web Journal*, 1(3), 77-84. http://www.cc.gatech.edu/gvu/user_surveys/papers/w3j.html

Lee, S. H., and Wang, C. J. (1997). Intelligent Hypermedia Learning System on the Distributed Environment. In *Proceedings of ED-MEDIA/ED-TELECOM'97, World Conference on Educational Multimedia/Hypermedia and on Educational Telecommunications* (pp. 625-630). ACCE.

León, H. G. B., Peñalvo, F. J. G., Rodriguez-Conde, M. J. (2008). Construction of Assessments with Double Adaptation Processes, *Innovative Techniques in Instruction Technology, E-learning*, Springer, pp. 156-160.

Leung, C.-K., Chang, H.-H., and Hau, K.-T. (2003) *Computerized adaptive testing: A comparison of three content balancing methods*. Journal of Technology, Learning, and Assessment 2(5) http://www.jtla.org

Lord, F. M. (1970). Some test theory for tailored testing. In W. H. Holtzman (Ed.) *Computer assisted instruction, testing and guidance* (pp. 139-183). New York: Harper and Row.

Microsoft CAT, 1999 *Microsoft Unveils Innovative Testing Technology to Simulate Work Environment*, http://www.microsoft.com/presspass/press/1999/jan99 /innovativepr.mspx

Sitthisak, O., Gilbert, L., and Davis, H. (2008). An evaluation of pedagogically informed parameterised questions for self assessment. *Learning, Media and Technology*, 33(3), 235-248.

Sitthisak, O., Gilbert, L., and Davis, H. C. (2007) Towards a competency model for adaptive assessment to support lifelong learning. *TEN Competence Workshop on Service Oriented Approaches and Lifelong Competence Development Infrastructures*, Manchester, UK.

Sosnovsky, S. (2004). Adaptive Navigation for Self-assessment quizzes. *Lecture Notes in Computer Science*, 3137, 365-371.

Thissen, D., and Mislevy, R. J. (2000). Testing Algorithms. In Wainer, H. (Ed.) *Computerized Adaptive Testing: A Primer*. Lawrence Erlbaum Associates, Mahwah, NJ.

Topic Maps (2000) Topic Maps, www.topicmaps.org

Tzanavari, A., Retalas, P., and Pastelis, S. (2004). Giving Adaptation Flexibility to Authors of Adaptive Assessments. *AH 2004*, 340-343.

van der Linden, W. J. and Glas, C.A.W. (Eds.). (2000). *Computerized adaptive testing: Theory and practice*. Kluwer, Boston, MA.

Wainer, H. (Ed.). (2000). *Computerized adaptive testing: A primer* (2nd edition). Mahwah, NJ: Lawrence Erlbaum.

Wainer, H., and Mislevy, R.J. (1990). Item response theory, calibration, and estimation. In Wainer, H. (Ed.) *Computerized Adaptive Testing: A Primer*. Mahwah, NJ: Lawrence Erlbaum Associates

Weber, G., and Brusilovsky, P. (2001). ELM-ART: An Adaptive Versatile System for Web-based Instruction. *International Journal of Artificial Intelligence in Education*, 12, 351-384.

Weber, G., and Specht, M. (1997). User Modeling and Adaptive Navigation Support in WWW-based Tutoring Systems. In *Proceedings of the 6th International Conference on User Modelling UM'97* Vienna, New York: Springer.

Wise, S. L., and Kingsbury, G. G. (2000). Practical issues in developing and maintaining a computerized adaptive testing program. *Psicologica*, 21, 135-155.

Huang, Y. M., Lin, Y. T., Cheng, S. C. (2009). An adaptive testing system for supporting versatile educational assessment, *Computers & Education*, 52 (1), 53-67.