An Adaptive Web Tool for Self-assessment using Lightweight User Profiles

Fotis Lazarinis[1], Vassilios S. Verykios[1] and Chris Panagiotakopoulos[2]
[1]School of Science and Technology, Hellenic Open University, Patras, Greece
[2]Department of Primary Education, University of Patras, Patras, Greece

Keywords: Assessment, Adaptive Educational Hypermedia, AEH, Reusability, IMS QTI, Topic Maps, Metadata.

Abstract: This paper presents an adaptive tool for self-assessments. The proposed system supports the selection of assessment items from an item bank based on a number of criteria such as the topic, the difficulty level of the items and a lightweight learner profile. For interoperability reasons, the assessment items are encoded using the IMS QTI standard and the topics are represented in Topic Maps XML. Items are included in the Topic Map as occurrences in one or more subtopics. The items are retrieved using parameterized XQuery scripts and they are adaptively presented to the user based on their knowledge level. Furthermore, some visual clues are associated to the items in the test that participants should attempt. The evaluation experiments showed that the tool supports more effectively self-assessment and motivates users to be more actively engaged.

1 INTRODUCTION

Formative assessment is defined as "the process used by teachers and students to recognise and respond to student learning in order to enhance that learning, during the learning" (Cowie and Bell, 1999). This broad definition allowed different forms of formative assessment to emerge. With advances in educational technology, terms like self, peer, collaborative, goal based assessment and the like are common. Of particular importance, is the type of formative assessment referred to as self-assessment, as students are always self-assessing, before exams or before handing in essays and reports. This kind of assessment is often informal and ad hoc but it is an important part of learning and therefore it should be treated more systematically (Boud, 1995).

Adaptive assessment refers to the ability of testing tools to adapt the testing process to the abilities or goals of learners. The most commonly applied adaptive test is CAT (Computer Adaptive Test) where the presentation of each item and the decision to finish the test are automatically and dynamically adapted to the answers of the examinees and therefore on their proficiency (Thissen and Mislevy, 2000). Alternative adaptive testing tools have also been proposed which focus on factors such as the competencies of the learners (Sitthisak et al., 2007) and their goals and current knowledge (Lazarinis et al., 2010).

In this work, we are interested in supporting adaptation in self-assessments in order to more effectively support the users' goals. The main aim of our work is to allow learners to adjust the testing material to their current goals and needs. In the current research design, learners are able to self-adapt the testing sequence to self-assess their knowledge. The process needs minimal input from the learners, such as their current knowledge on their targeted topics and their current testing preferences (e.g., the difficulty of the testing items). They can also re-adapt their goals during the testing procedure to adjust it on their evolving goals.

2 RELATED WORKS

Most of the computerized adaptive testing tools are based on the Item Response Theory and they estimate the knowledge of each student with a shorter number of queries tailored to the performance of each test participant (van der Linden and Glas, 2000). A criticism to this approach is that they are based solely on the performance of the students, which limits their use for alternative educational purposes (Wise and Kingsbury, 2000; Wainer, 2000).

Lazarinis F., S. Verykios V. and Panagiotakopoulos C..
An Adaptive Web Tool for Self-assessment using Lightweight User Profiles.
DOI: 10.5220/0005403700140023
In *Proceedings of the 7th International Conference on Computer Supported Education* (CSEDU-2015), pages 14-23
ISBN: 978-989-758-108-3

Therefore, a variety of alternative approaches in adaptive assessment tools have been proposed. QuizPACK (Brusilovsky and Sosnovsky, 2005) and QuizGuide (Sosnovsky, 2004) support self-assessment of programming knowledge with the aid of Web-based individualized dynamic parameterized quizzes and adaptive annotation support. Multiple versions of the same queries are offered to learners who can see the right answers and try the same question again but with different parameters. The tools described in these papers are domain dependent and are basically possible in domains such as mathematics, physics and programming.

In another proposal, a set of competencies are defined at the beginning while the next assessment stages rely on the competencies an individual possesses (Sitthisak et al., 2007). The competencies rely on parameterized attributes and thus they can be modified for different domains. This work is extended in a later work (Sitthisak et al., 2008) where the authors present a tool for automatically creating a number of questions for a required competency, based on the associations of the questions to various competencies.

A method for evaluating learning achievement and providing personalized feedback of remedial suggestion and instruction for learners is presented in (Yi-Ting, 2012). First learners' test results are calculated in terms of accuracy rate, test difficulty, confidence level, and length of answer time. Personalized feedback for learners based on concept map with cognitive taxonomy is provided.

Decision trees and rules are used for adapting the testing procedure in another e-learning environment (Šerbec et al., 2011). The adaptation of the testing procedure relies on the performance, the current knowledge of test participants, on the goals of educators and on the properties of knowledge shown by participants. Collaborative annotating and data mining are employed into formative assessments to develop an annotation-sharing and intelligent formative assessment system as an auxiliary Web learning tool (Lin and Lai, 2014).

In one of our previous works we developed an adaptive 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 (Lazarinis et al.). Educators outline adaptive assessments by using IMS QTI (2006) encoded items and customizable rules. IMS QTI defines a standard format for the representation of assessment content and results. Test creators associate specific conditions at various points of the testing sequence which, if they are met, change the testing path and adapt it to the characteristics of the individual user. The learners' data are encoded in IMS LIP (2005) standardised structures for learner profiles and include data about the knowledge level of learners per topic and their goals and preferences. This research proposal supports mainly the educational strategies of educators.

Individualized skill assessment has been proposed for digital learning games (Augustin et al., 2011). A new problem to be presented to users is based on their previous problem solving process and their competence state. An adaptive assessment system with visual feedback is described in (Silva and Restivo, 2012). Adaptive test generation based on user profiling is utilized in a personalized intelligent online learning system (Jadhav, Rizwan and Nehete, 2013). An adaptive fuzzy ontology for student learning assessment applied to mathematics is presented in (Lee et al., 2013). The purpose of the study is to understand the weaknesses of the students. In a web-based system for self-assessment, learners can freely select the tests or navigate through them in a linear mode (Antal and, Koncz, 2011).

The above studies show that there is a growing body of researchers interested in providing adaptivity features to assessment systems to support the aims of the users and to diagnose their knowledge and difficulties. The applied adaptive techniques are based on different factors and not only on their performance but they either require detailed student profiles or are domain specific. In the current study we are developing and evaluating an adaptive tool for self-assessments using limited user information applicable to various domains.

3 SYSTEM DESCRIPTION

Self-assessment has been shown to support student learning (Taras, 2010). The main goal of the current work is to provide a flexible environment for self-assessment, where the test participants can regulate the testing process based on their current goals. Adapting the testing process to their current learning goals is expected to have multiple benefits to the knowledge, the self-efficacy and self-esteem of learners. Further aims of the proposed design are to be domain independent and to require minimal learner information to be used in various learning situations and computer environments. Most of the tools presented in the previous sections require detailed learner profiles to be effective. This makes them inflexible as the required data may not be available or learners may not be willing to share extensive personal information.

In order to achieve these goals, we designed a modular adaptive application consisting of an authoring environment for developing IMS QTI compliant items associated with specific topics and a run time module where test participants can:

- select one or more topics;
- define their knowledge level and the level of completed education on the selected topics;
- define the characteristics of the assessment items they want to try;
- define the number of items and finally execute the assessment.

Based on the knowledge level and the learner's performance, a number of inferences about the knowledge of the test participants in the specific topics are possible.

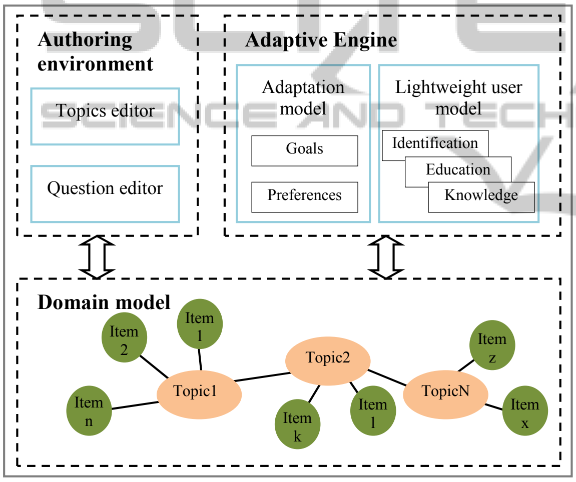

Figure 1: Components of the adaptive testing tool.

Typically, adaptive e-learning tools consist of the domain model, the user model, the adaptation model and the adaptive engine (De Bra et al., 2004). Following this paradigm, our proposed adaptive information system consists of a domain model which consists of the topics, their associations and the assessment items (Figure 1). The user model consists of some identification data (name, email), education (e.g., high school student) and the knowledge level in some topics. The adaptation model is a collection of rules that define how the adaptation must be performed. In our case, the adaptation is realized by letting users define the criteria about the items which would be presented to them. The adaptive engine is the module which retrieves the relevant items and supports the execution of the assessment and finally presents the results to the user.

3.1 Topics and Assessment Items

The authoring environment supports the development of topic networks and assessment items (Figure 2). Assessment items are encoded in IMS QTI and for each item several metadata can be defined, e.g., the educational level, the difficulty level (easy, medium, and difficult), feedback, etc. (Figure 3). Some of these metadata are encoded in IEEE LOM (2002) (i.e. Learning Object Metadata) under the <general> and <educational> elements. These two standards are packaged using an IMS manifest (imsmanifest.xml) file which includes a reference to the respective IMS QTI XML file and the necessary metadata under the <imsmd:lom> element. Since the assessment data are encoded using standardized XML structures, compliant items from external sources may be utilized, extending the existing item bank. Further, any IMS QTI compliant editor could be used for authoring assessment items complying with the latest versions 2.x. of the standard.

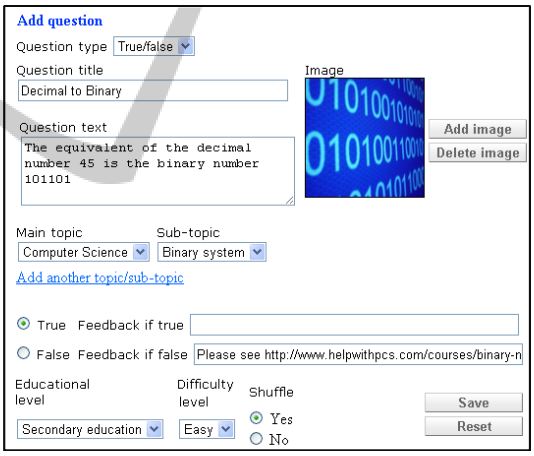

Figure 2: Assessment item editor.

The next part of the domain model concerns the topics (e.g., physics, computer science, literature etc), their subtopics and the associations between subtopics and the assessment items. The information about the topics is encoded using the XML Topics Maps (XTM) (2000). This is a standardized encoding scheme for representing the structure of information resources used to define topics and associations (relationships) between topics. Subtopics are associated to one or more topics using the <instanceOf> element of XTM. The <association> element is used to associate topics and to create topic classes. For example, physics and chemistry are grouped under the Science association.

Assessment items are represented as occurrences, using the <occurrence> element of XTM, in one or more subtopics and thus these items are implicitly associated with the parental topics. The XML files of the manifest files packaging the assessment items and their metadata are referenced into the <occurrence> element using the <resourceRef> sub-element (see Figure 4). As in the case of assessment items, the topics maps can be edited using any compliant tool, e.g., Ontopia.

Educators are able to add new assessment items and associate them with one or more of the subtopics. Initially, the main classes of topics are represented in tree view. The test creators can expand this tree to locate the appropriate subtopic and associate it with the edited assessment item. Representing the information by using standardized semantic technologies, increases the sharing and reusability of information facilitating the integration of compliant resources.

```
<assessmentItem … identifier="choice"
title="Decimal to Binary">
<responseDeclaration identifier="q_1"
cardinality="single" >
<correctResponse><value>ChoiceA</value>
</correctResponse>
</responseDeclaration>
 <itemBody>
      <p><img
src="images/q_1_1.png"></p>
     <choiceInteraction
responseIdentifier="RESPONSE"
shuffle="true">
    <prompt>The equivalent of the
decimal number 45 is the binary number
101101</prompt>
    <simpleChoice
identifier="ChoiceA">True</simpleChoice
>
    <simpleChoice
identifier="ChoiceB">False
      <feedbackInline
outcomeIdentifier="FEEDBACK"
identifier="q_1" showHide="show">Please
see
http://www.helpwithpcs.com/courses/bina
ry-numbers.htm#decimal-to-binary-
conversion
     </feedbackInline>
    </simpleChoice>
</choiceInteraction>
  </itemBody>
</assessmentItem>
```

Figure 3: Question encoded in IMS QTI.

```
<topic id="Binary-System">
  <occurrence id="q_1">
    <instanceOf>
       <topicRef xlink:href="#xml-
version"/>
    </instanceOf>
    <resourceRef
xlink:href="imsmanifest_q_1.xml"/>
  </occurrence>
</topic>
```

Figure 4: Topic Map occurrence.

3.2 Adaptation Process

The adaptation process is realized during the execution phase. First, users have to provide some information about themselves, in order to be identified into the system. The minimum information required is a name or nickname and an email to communicate the test results. Then they need to inform the system about their goals and preferences. That is, they have to define the topics they wish to be assessed on, the educational level of the assessment items, the difficulty level, and the number of questions. During the execution, they can also get feedback on each question. If they wish they can define their educational level and their estimated knowledge on the selected topics.

The identification data, the educational level and the estimated knowledge on the selected topics compose a lightweight user profile which is used in the adaptation of the content and is active only during the assessment, although these data could be stored in a user profile with the consensus of the users for exploitation in future self-assessment. The selected topics as well as the defined educational and the difficulty level of the assessment items comprise the *adaptation model*.

Test participants can select assessment items by defining one of the following options:

i. The difficulty level and/or the educational level of the questions: Since questions are classified as easy, medium or difficult, learners can select assessment items based on their difficulty. They can select questions that equal or are above or below a specific difficulty level, e.g., “show only difficult questions”.

ii. Questions based on the learner’s knowledge level: students can select questions that match or are above or below their knowledge level. Easy, medium and difficult questions match to low, good and very good knowledge level. So if a student, for example, has a “good” knowledge level in a topic, then s/he can form rules like “show questions that match or exceed my knowledge level” and the system will retrieve testing items of medium or higher difficulty.

iii. Questions based on the learner's educational level: with this option, questions are selected based on the educational level of the question. Students are able to form rules like "show questions that match my educational level".

Alternatively a learner can let the system decide the sequence of questions based on the data the student inputs about her/his knowledge and educational level. In that case, the application retrieves questions that match the learner's educational level and are sorted according to their difficulty level.

As we can see in figure 5, the completion of this information is a straightforward process. Users have to complete a single form by typing or selecting the appropriate options. At any given point during the test they can change the input options to retrieve a different set of assessment items.

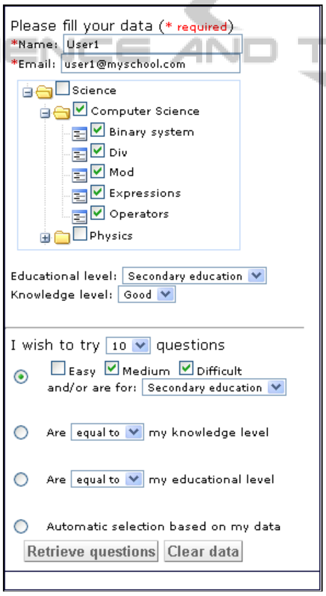

Figure 5: Adaptive selection of questions.

The adaptive engine is based on parameterized nested XQuery scripts for querying and processing the topic maps and the packaged QTI items which operate on the XML of the topic maps and then on the packaged assessment items. These queries select the matching assessment items. The scripts take the user defined adaptation options, e.g., the desired topics, as input data. Then a list of assessment items is formed.

The presented items are grouped based on the subtopic they relate to. If there are remaining questions in a subtopic, these are grouped at the end of the assessment items under a "Similar questions" button (Figure 6).

As seen in figure 6, users are presented with lists of assessment items which match their input options. Items are grouped based on the subtopic they relate to. Further, the links are adaptively presented and annotated (Brusilovsky, 2001) based on the previous knowledge of the test participants as it was stated at the beginning of the test and the current knowledge as it is estimated by the system. Adaptive presentation means that the groups of items that have higher difficulty level than the user's defined knowledge level, are presented first. Adaptive annotation refers to the attachment of visual clues to items that the system believes a user has to attempt. One such clue is the red exclamation mark in front of an item which in essences prompts users to attempt these items first. Further, if users fail one of the questions of lower difficulty level than the user's defined knowledge level in a subtopic, then the rest of the questions in this subtopic are emboldened to help them understand that they need to attempt all the related questions.

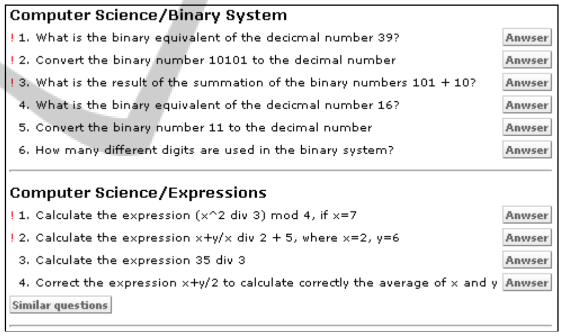

Figure 6: Presentation of assessment items to a student with average knowledge on the selected topics. Questions with higher difficulty are preceded by a red exclamation mark.

The new knowledge level per topic is based on the average test score on the specific topic. If the average is below 50% then the knowledge level is set to "low"; an average between 50% and below 75% results in a knowledge level set to "good"; scores 75% or higher are treated as a "very good" knowledge level. The same process is applicable to the estimation level of each subtopic. At the end of the test the results per topic and subtopic are presented and the estimated knowledge level and the erroneously items with the available feedback are given to the system.

In case a user provided his/her initial knowledge on the topics then the system presents the initial knowledge level and the estimated knowledge.

Finally, the results are emailed to the user for future reference.

Further, statistics per question are stored in a separate repository in order to be used by the system and the test creators. For each user session the user's initial knowledge level, the final knowledge level and the result (correctly answered/wrongly answered) for each attempted question is stored. This information will be used to help educators revise the classification and the phrasing of questions and support more adaptation options in the future based on automatic question classification.

4 EVALUATION

The proposed system supports the automatic selection of assessment items from an item bank based on user defined criteria. The retrieved items are sorted based on their topics, their difficulty level and the learners' current knowledge level. The system aims at being a flexible environment, supporting various adaptation techniques which produce a list of assessment items for self-assessment.

To assess its significance, different evaluation experiments, which will test the system's usefulness, the help and motivation provided to students, need to be carried out

The questions of the current initial evaluation were:

a. To understand if the system motivates students to be more actively engaged in the process of self-assessment.
b. To evaluate the ability of the system to better adapt to the needs of the learners.
c. To measure the potential improvement on the performance of the learners in summative assessments.

The experiment was carried out with the help of 106 high school students (aged 17 to 18) who attend the last two final classes of high school. Due to the increased number of participants, data gathering was administered in different periods during May 2014 and November 2014. The participants provided an estimation of their knowledge levels prior to the evaluation of the system, to be able to uniformly distribute the learners into two groups. We divided the students in two groups ensuring that students of different knowledge levels in the subject of "introductory algorithmic concepts" are included in both groups. Students of similar knowledge levels were randomly assigned to one of the two groups. Then we batch converted 200 questions (true-false, single and multiple choice, fill-in-the-gap) to IMS QTI XML and assigned to their respective subtopics (e.g., Div operator) of the "Introduction to Algorithms" topic. The educational level was "Secondary Education" and for each question we also included their difficulty level and the correct response.

The first group of students used a non-adaptive version of the system and consisted of 49 students. The group consisted of 8 students with a low knowledge level, 25 students with knowledge on the topics of the test and 16 students with very good knowledge level. Each student could decide the number of questions s/he wanted to try and then the respective number of questions was randomly selected from the item bank. The students did not have options like "Similar questions" or "More questions". They could of course re-run the application at the end of an assessment, should they wished.

The second group of 57 students used the adaptive version of the system with the options described in the previous sections, but we made all of the features optional to see whether students would actually use them. This group included 10 students of low knowledge in the topics of the tests, 29 students of good knowledge and 18 students of high knowledge level. The students' knowledge levels are uniformly distributed among the two groups of students.

All the students of both groups were informed that they had to study for a regular summative test at the end of the trimester. So, before the evaluation experiment, they were informed that they had to study the appropriate learning material and then to use the self-assessment tool for up to 45 minutes in order to self-assess their knowledge. The activities of the students were recorded into log files to be studied later. Also, during the manual analysis of the log files, a short focused interview was conducted with each student separately.

4.1 Qualitative Analysis of the Results

The qualitative study of the log files pointed out that the first group of participants selected 10-20 questions (mean 14.57, median 15). Table 1 shows the number of students and the respective number of selected questions. None of these students re-ran the system to try new questions. When asked, the students argued that it would be a tedious process to restart the test or they have not thought of that possibility.

Table 1: Number of questions selected in the non-adaptive self-assessments.

No of students	No of questions selected by each student
7	10
11	12
8	14
6	15
5	16
5	18
7	20
7	10
Avg. number of questions per student: 14.57	

Table 2: Number of initially selected questions in the adaptive self-assessments.

No of students	No of questions selected by each student
2	8
8	10
9	12
8	14
7	15
6	16
9	18
8	20
Avg. number of questions per student: 14.72	

Table 2 concerns the second group. The second column shows the number of questions that were initially selected by the students (mean 14.72, median 14). We observe that the mean number of initially selected questions is similar in both students groups. Even so, the students of the second group finally attempted more questions than the number of items that they initially defined. Through the utilization of the “Similar questions” or “More questions” buttons, more problems were shown to the learners. The average number of questions finally answered by the students of the second group increased to 19.40 (median 18). This increase in testing items varies from 10% to 90%. For example, one student of group 2 had initially selected to answer 10 questions and s/he finally answered 19 items.

We asked each student to explain why they tried more questions. 51 of the students of the second group replied that they used the “Similar questions” option in some subtopics and answered more questions than their initial intentions. 6 students of the second group the button used the “More questions” which appeared at the end of the list of the assessment items. 10 students had initially defined a small number of questions and therefore the button “More questions” appeared at the end of the list of the assessment items and 6 of them used it. In all the other cases the button “Similar questions” appeared in one or more subtopics. All the students argued that these options encouraged them to try more questions.

These results are strong indications that our tool motivates the students to be more actively engaged in the process of self-assessment by answering more questions than their initial intents.

The next step is associated with the second aim of the evaluation. As said, all the adaptation options were made optional for the second group of students. The students were also informed that they are not obliged to use any of the available options to ensure that none of the participants will reluctantly select some of the rules. At the end of the self-assessment of the group 2 students, we recorded the options they used. First, we observed that all the students used one of the available adaptation options. This result in conjunction with the usage of the “Similar questions” and “More questions” during the test by the participants, are positive signs towards the ability of the tool to better adapt to the needs of the learners.

41 of the 57 participants used the adaptation options which options related with the difficulty of the questions i.e. "try questions of specific difficulty level". 12 students used the adaptive option related to their knowledge level and the rest 4 students let the system adapt the process based on their user profiles. More tests are however needed, to understand the usefulness of each adaption option and to realize if more options are necessary.

In the next stage of the evaluation experiment, a short focused interview was conducted with each participant. According to the answers, students of both groups found their version of the system easy to use. Further, the students of the second group considered as very important the fact that they could adjust the features of the testing items. Another positive aspect of the system is that the most difficult questions appeared first in the list of assessment items and with a red exclamation mark. Technically speaking, these questions are those that are of a higher difficulty level than their knowledge level.

4.2 Post Evaluation Assessment

After the utilization of the system from both groups, students had to take a non-adaptive summative 20-question e-test on the same topics. The 20 questions were not included in the item bank used in the self-assessment, but they concerned the same topics. The aim of the summative test was to see whether there is any difference in the performance of the learners after the usage of the adaptive version of the tool. This aim is associated with the last question of the evaluation.

After the summative test, 42 out of 49 students of the first group achieved scores in accordance with the knowledge level they were classified at the self-assessment test. 5 students had a worse performance in the summative test in comparison with the self-estimated knowledge level. 2 students of the first group classified as having good knowledge, i.e. higher knowledge than the learner estimated knowledge level. The increase or drop of the performance with respect to the users' estimated knowledge level may be due to the increased or reduced number of questions they tried during the self-assessment test. Or we could suppose that the initially user estimated knowledge level was inaccurate. In general it is risky to attribute the improvement to a specific reason, without conducting extensive evaluations.

48 students of the second group had the same classified in the same knowledge level as it was estimated in the self-assessment. The remaining 9 students of the second group had a better performance than their initial user provided knowledge level. We asked these students why they believed they were classified in a higher class in the summative test. They mentioned that they tried many questions in the self-assessment and therefore the testing items concerned similar concepts. This belief is positive towards our research proposal. As it motivates students to be more actively engaged in the process of self-assessment it is reasonably expected that the students will eventually perform better in summative assessments on similar topics. But as previously noted, this may be due to other factors, i.e. an inaccurate user estimated knowledge level.

Table 3 shows the scores in the summative tests. The mean score is higher for students who used the adaptive version of the system and attempted more questions of the difficulty level they defined or exceed their knowledge level. Running an independent two-tailed t-test on the results with a null-hypothesis that 'there is a no statistical difference between the two means', we come up with $p=0.041<5\%$. This means that we can reject the null-hypothesis with high confidence as the probability of being wrong is less than 5%. In any case, the main purpose of this first evaluation experiment was to qualitatively estimate the usefulness of the tool and to understand whether there are positive indications for our research direction. Through a longitudinal study with different student populations and different question items falling under various thematic areas would strengthen our findings.

Table 3: Scores in summative assessment.

No of students	Score of group 1 students	No of students	Score of group 2 students
9	< 50%	5	< 50%
27	>=50% and <75%	30	>=50% and <75%
13	>=75%	22	>=75%
Avg. score: 13.26 (65%)		*Avg. score: 15.32 (77%)*	

The main conclusion of the evaluation experiment is that learners find our tool useful and the available options motivate them to be engaged more actively in the process of self-assessing their knowledge. Although more than 120 different items were answered during the test, more evaluations are necessary to understand the long term effects on knowledge improvement and learner's motivation.

5 DISCUSSION AND FUTURE WORK

In the previous sections a system for self-assessment was presented which allows learners to define the criteria for selecting the assessment items. Users provide a lightweight profile consisting of an expression of their previous education and goals and the system selects the most appropriate items from an item bank. The questions of the item bank are associated with specific topics, educational and difficulty level and are represented in standardized XML structures which make the utilization of external resources easier. The system orders the retrieved set of items based on their difficulty level and the user provided knowledge level. Visual clues are attached to each question based on the initial learner knowledge level and on the user performance. Test participants are able to access more assessment items, similar to the presented ones, for additional testing of their knowledge.

The initial evaluation showed that the system is useful and that the students are engaged more actively based on the available options and adaptive features of the system. Students explored more testing items than originally selected and also achieved demonstrably higher levels of learning.

Several extensions are possible in such a system. First, another topic map could be developed linking assessment items and concepts to specific lessons or certifications so that students can adapt their selection process accordingly. Information about the item creator could be also utilized to the benefit of the students, especially within a specific institute. Further, the capabilities of the item bank should be extended, with additional categories of assessment items which will be more interactive, e.g., java applets or flash animations. Apart from the current feedback, specific links to external sources or fragments of learning material could be associated to assessment items or topics and be presented to the users to help them study the materials with the greatest difficulty. Some of these improvements are already under development along with the design of new evaluation experiments.

ACKNOWLEDGEMENTS

The work is partially supported by the Google CS4HS program.

REFERENCES

Antal, M., Koncz, S., 2011. Student modeling for a web-based self-assessment system. *Expert Systems with Applications*, 38(6), 6492–6497.

Augustin, T. , Hockemeyer, C., Kickmeier-Rust, M., Albert, D., 2011. Individualized skill assessment in digital learning games: basic definitions and mathematical formalism. *IEEE Transactions on Learning Technologies*, 4(2), 138-148.

Boud, D. 1995. *Enhancing learning through self-assessment*, Routledge.

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*, 5(3), http://doi.acm.org/10.1145/1163405.1163411.

Cowie, B., Bell, B. 1999. A model of formative assessment in science education. *Assessment in Education*, 6, 101–116.

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

IEEE LOM, 2002. *Learning Object Metadata*, http://ltsc.ieee.org/wg12/

IMS LIP, 2005. *Learner Information Package*, http://www.imsglobal.org/profiles.

IMS QTI, 2006. *Question and test interoperability*. http://www.imsglobal.org/question.

Jadhav, M., Rizwan, S., Nehete, A., 2013. User profiling based adaptive test generation and assessment in e - learning system. *3rd IEEE International Advance Computing Conference*, pp. 1425-1430 DOI: 10.1109/IAdCC.2013.6514436.

Lazarinis, F., Green, S., Pearson, E. 2010. Creating personalized assessments based on learner knowledge and objectives in a hypermedia Web testing application. *Computers and Education*, 55(4), 1732-1743.

Lee, C. S., Wang, M. H. Chen, I. H., Lin, S. W., Hung, P. H., 2013. Adaptive fuzzy ontology for student assessment. *9th International Conference on Information, Communications and Signal Processing*, ICICS, pp. 1-5, DOI: 0.1109/ICICS.2013.6782880.

Lin, J.-W., Lai, Y.-C., 2014. Using collaborative annotating and data mining on formative assessments to enhance learning efficiency. *Computer Applications in Engineering Education*. 22(2), 364-374. doi: 10.1002/cae.20561.

Šerbec, I.N., Žerovnik, A., Rugelj, J., 2011. Adaptive assessment based on decision trees and decision rules. *CSEDU 2011 - Proceedings of the 3rd International Conference on Computer Supported Education*, pp. 473-479.

Silva, J. F., Restivo, F., J., 2012. An adaptive assessment system with knowledge representation and visual feedback. *15th International Conference on Interactive Collaborative Learning*, pp. 1-4, IEEE Digital Library DOI: 10.1109/ICL.2012.6402117.

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.

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

Sosnovsky, S., 2004. Adaptive Navigation for Self-assessment quizzes. *LCNS vol.* 3137, pp. 365-371.

Taras, M., 2010. Student self-assessment: processes and consequences. *Teaching in Higher Education*, 15(2), 199-209.

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.

van der Linden, W. J., 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.

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

Yi-Ting, K., Yu-Shih, L., Chih-Ping, C., 2012. A multi-factor fuzzy inference and concept map approach for developing diagnostic and adaptive remedial learning systems. *Procedia - Social and Behavioral Sciences*, 64, pp. 65–74.