Evaluation of Topic-based Adaptation and Student Modeling in QuizGuide

Sergey Sosnovsky[1], Peter Brusilovsky[2]
[1] German Research Center for Artificial Intelligence (DFKI),
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
phone: +49 (0) 681 / 85775-5367
fax: +49 (0) 681 / 85775-1053
sosnovsky@gmail.com
[2] University of Pittsburgh, School of Information Sciences,
135, N. Bellefield Ave., Pittsburgh, PA, 15260, USA
phone: +1 (412) 624 9404
fax: +1 (412) 624 2788
peterb@pitt.edu

Corresponding Author:
Sergey Sosnovsky
E: sosnovsky@gmail.com
T: +49.681.85775.5367
F: +49.681.85775.1053

Abstract:

This paper presents an in-depth analysis of a nonconventional *topic-based* personalization approach for adaptive educational systems (AES) that we have explored for a number of years in the context of university programming courses. With this approach both student modeling and adaptation are based on coarse-grained knowledge units that we called topics. Our motivation for the topic-based personalization was to enhance AES transparency for both teachers and students by utilizing typical topic-based course structures as the foundation for designing all aspects of an AES from the domain model to the end-user interface. We illustrate the details of the topic-based personalization technology, with the help of the Web-based educational service QuizGuide – the first system to implement it. QuizGuide applies the topic-based personalization to guide students to the right learning material in the context of an undergraduate C programming course. While having a number of architectural and practical advantages, the suggested coarse-grained personalization approach deviates from the common practices toward knowledge modeling in AES. Therefore, we believed, that several aspects of QuizGuide required a detailed evaluation – from modeling accuracy to the effectiveness of adaptation. How can one evaluate the soundness of a user modeling (UM) approach, rather than a specific personalized system? The paper discusses how this new student modeling approach can be evaluated, and presents our attempts to evaluate it from multiple different prospects. The evaluation of QuizGuide across several consecutive semesters demonstrates that, although topics do not always support precise UM, they can provide a basis for successful personalization in AESs.

Keywords:

Adaptive Hypermedia, Student Modeling, Adaptive Educational System, Adaptive Navigation Support, Adaptive Link Annotation, Topic-based Adaptation, Topic-based User Modeling, Adaptive System Evaluation, Layered Evaluation

The first page should include the following declaration: "This paper or a similar version is not currently under review by a journal or conference, nor will it be submitted to such within the next three months. This paper is free of plagiarism or self-plagiarism as defined in Springer's Policy on Publishing Integrity.

Evaluation of Topic-based Adaptation and Student Modeling in QuizGuide

Sergey Sosnovsky[1], Peter Brusilovsky[2]

[1] German Research Center for Artificial Intelligence (DFKI),
Stuhlsatzenhausweg 3, D-66123 Saarbrücken, Germany
phone: +49 (0) 681 / 85775-5367
fax: +49 (0) 681 / 85775-1053
sosnovsky@gmail.com

[2] University of Pittsburgh, School of Information Sciences,
135, N. Bellefield Ave., Pittsburgh, PA, 15260, USA
phone: +1 (412) 624 9404
fax: +1 (412) 624 2788
peterb@pitt.edu

Abstract:

This paper presents an in-depth analysis of a nonconventional *topic-based* personalization approach for adaptive educational systems (AES) that we have explored for a number of years in the context of university programming courses. With this approach both student modeling and adaptation are based on coarse-grained knowledge units that we called topics. Our motivation for the topic-based personalization was to enhance AES transparency for both teachers and students by utilizing typical topic-based course structures as the foundation for designing all aspects of an AES from the domain model to the end-user interface. We illustrate the details of the topic-based personalization technology, with the help of the Web-based educational service QuizGuide – the first system to implement it. QuizGuide applies the topic-based personalization to guide students to the right learning material in the context of an undergraduate C programming course. While having a number of architectural and practical advantages, the suggested coarse-grained personalization approach deviates from the common practices toward knowledge modeling in AES. Therefore, we believed, that several aspects of QuizGuide required a detailed evaluation – from modeling accuracy to the effectiveness of adaptation. How can one evaluate the soundness of a user modeling (UM) approach, rather than a specific personalized system? The paper discusses how this new student modeling approach can be evaluated, and presents our attempts to evaluate it from multiple different prospects. The evaluation of QuizGuide across several consecutive semesters demonstrates that, although topics do not always support precise UM, they can provide a basis for successful personalization in AESs.

Keywords:

Adaptive Hypermedia, Student Modeling, Adaptive Educational System, Adaptive Navigation Support, Adaptive Link Annotation, Topic-based Adaptation, Topic-based User Modeling, Adaptive System Evaluation, Layered Evaluation

1. Introduction

Web-based education has proven to be a very productive application area for adaptive hypermedia (Brusilovsky 1996, 2001; Knutov et al. 2009). Many adaptive Web-based educational systems explore a wide range of adaptive hypermedia techniques (such as page content or media adaptation, link annotation, hiding, or generation) and attempt to use various characteristics of a student (goals, knowledge, learning style, etc.) as sources of adaptation. Many cases of success have been reported. According to the results of published studies, the use of adaptive hypermedia helped to improve student learning, enabled students to more quickly achieve their learning goals, reduced navigational overhead, and increased student satisfaction (Olston,Chi 2003; Kavcic 2004; Davidovic et al. 2003; Brusilovsky,Eklund 1998; Hothi et al. 2000; Triantafillou et al. 2004). Despite all the demonstrated benefits, the use of adaptive educational hypermedia systems in practical education is still rather modest. Instead, the global Web-based education market is covered by far less sophisticated learning management systems (LMS), which offer minimal (or no) personalization.

The low use of adaptive hypermedia in everyday Web-based education has been often attributed to the complexity of authoring adaptive educational applications (Brusilovsky 2003) and the lack of proper authoring tools. In response to this recognized need, the adaptive hypermedia research community developed a range of authoring tools, frameworks, and toolkits (De Bra et al. 2003; Weber et al. 2001; Cristea,Aroyo 2002; Brusilovsky et al. 1998; Carmona et al. 2002; Vassileva 1998; Aroyo,Dicheva 2001; De Bra et al. 2013; Specht,Oppermann 1998). These tools certainly helped to reduce the authoring barrier for adaptive educational hypermedia, yet they failed to produce a visible impact on everyday practice. While a number of adaptive courses and applications were produced using such tools, they were predominantly authored by academics either for their own courses or in collaboration with instructors. However, these tools have not been embraced by regular instructors for their authoring needs in the same way that they adopted traditional learning management systems.

Our experience with practical use of one of these authoring tools – InterBook (Brusilovsky et al. 1998) – has demonstrated that the source of problems is not just the technical complexity of the authoring process (which can be dramatically decreased by the authoring tool), but also its conceptual complexity. This complexity is caused by the dominant approach to knowledge and content structuring in adaptive educational hypermedia – *concept-based adaptation*. It assumes relatively fine-grained domain models with dozens to hundreds of concepts (often connected by links) and expects each educational content element to be indexed with several domain concepts. The concept-based approach is known to be very powerful: adaptive hypermedia systems based on this approach were able to achieve very sophisticated and precise adaptation. Not surprisingly, nearly all known authoring tools focus on supporting concept-based adaptive hypermedia. While it helps teams with adaptive educational hypermedia expertise to produce strong courses with sophisticated adaptation, we believe that might make the authoring process too complicated for regular instructors. In our own collaboration with instructors, we have observed that all components of the concept-based authoring process (creating fine-grained concept-level domain models, linking concepts to each other, and indexing content with concepts) are difficult procedures for instructors to comprehend and master. They think about the learning domain in terms of more coarse-grained *topics* structuring the curricula of their courses and are used to *classifying* each element of educational content as belonging to a certain topic, rather than *indexing* it with multiple concepts (Brusilovsky 2012).

The barrier between the complexity of modern concept-based authoring of adaptive hypermedia and the instructors' topic-oriented thinking about the domain could be resolved in several ways: (1) keep authoring in the hands of professionals while employing instructors as domain experts rather than authors; (2) increase the level of intelligence of authoring tools to automatically perform domain concept elicitation and content indexing; or (3) use to a less sophisticated approach of domain modeling and content structuring, which is closer to teachers' "native" understanding of the domain. While the first approach emerged as a popular practice and the second one provoked a stream of interesting research (Lawless et al. 2008; Smith,Blandford 2003; Specht et al. 2002; Wang,Taylor 2007; Hatala et al. 2009; Brusilovsky et al. 2006; Brusilovsky et al. 2005b; Sosnovsky et al. 2004; Sosnovsky et al. 2012; Apted et al. 2004), this paper focuses on our experience with less-investigated third approach.

Over the last several years, we have explored a simpler teacher-oriented approach to knowledge structuring, user modeling (UM) and adaptation in educational adaptive hypermedia, which we call the *topic-based approach*. The goal of our research has been to investigate the extent to which simplified instructor-oriented coarse-grained domain conceptualization can produce useful and efficient adaptive educational hypermedia. To explore the properties of this approach, we developed a topic-based adaptive hypermedia service QuizGuide, which provides personalized access to self-assessment learning material by guiding students to the most appropriate quizzes. The original version of QuizGuide was developed for the C programming domain (Brusilovsky,Sosnovsky 2005a). The success of this version encouraged us to implement QuizGuide for two more domains: SQL (Sosnovsky et al. 2008) and Java programming (Hsiao et al. 2010). By now, approximately 2000 students have used these three versions of the system in more than 90 courses. Altogether, the students answered more than 50,000 questions while receiving navigation support from QuizGuide.

This paper reports the results of our work on topic-based UM and adaptation, and presents an extensive evaluation of the approach. Section 2 introduces the topic-based approach and its specific application for C programming in QuizGuide. Section 3 explains important details of the implementation of the original QuizGuide system. Section 4 presents the evaluation experiment, including the context and design of the studies, the subject population and the collected data. It also overviews the range of methods we used for evaluating various aspects of QuizGuide. The results of analyses conducted according to these methods are reported in the next four sections. This part is the centerpiece of the paper and we believe that its value goes beyond an evaluation of a specific system or approach, but can be used as a reference point by other researchers examining their adaptive e-Learning solutions.

Section 5 provides a comprehensive overview of the effects QuizGuide has on a range of learning parameters, including knowledge gain, motivation and performance within the system. Section 6 uses student log data to thoroughly evaluate multiple layers of QuizGuide personalization approach. It starts by examining the quality of the designed topic-based domain model and the precision of knowledge assessment supported by an average topic. It continues with the student modeling mechanism employed in QuizGuide and compares its against Bayesian Knowledge Tracing (Corbett,Anderson 1995) in terms of predictive validity. Finally, it evaluates the impact of adaptive annotations generated by QuizGuide on student behavior. Section 7 takes a deeper look into the global navigational patterns shown by students and outlines the differences in students' self-regulated work with quizzes with and without adaptation. Section 8 reports the results of the students' evaluation of the system based on a set of post-course questionnaires.

Section 9 presents a brief overview of seminal work on student modeling in adaptive educational systems (AES). Section 10 discussed the main results of the paper and the prospects of the topic-based approach. Finally, Section 11 indicates potential directions for future work and concludes the paper. Altogether, this paper attempts to attract attention to the topic-based approach as a possible alternative for organizing adaptive hypermedia systems, offer a practical example of its implementation, and provide empirical evidence that highlights strong and weak sides of this approach.

2. Topic-based approach to knowledge modeling and adaptation

2.1. TOPIC-BASED CONTENT MODELING

The complexity of adaptive content creation primarily originates in the complexity of the domain models used in modern AESs. The more detailed and precise the modeling is, the more accurately the system can assess student knowledge; therefore, the more effectively it can potentially adapt its content to the individual student. However, the important questions are: Where does the "golden mean" lie? What is the best tradeoff between the model precision and model complexity, and between the effectiveness of adaptation and the ease of development? Can we reduce the authoring effort without jeopardizing the quality of adaptation?

Our collaboration with instructors interested in developing adaptive hypermedia content encouraged us to explore a minimalist *topic-based* approach toward domain modeling and content structuring. Its key idea is subdividing the subject to be taught into coarse-grained domain units (topics) and classifying each piece of educational content as belonging to exactly one of these topics. With this approach, the number of domain knowledge elements remains easy to manage, and the procedure of domain and content modeling, essentially, replicates the process of course design that many teachers follow. Separating the body of learning material into topics (or lectures) is a natural task for a classroom teacher preparing a course. It is also very common for a teacher to assign content (instructional texts, exercises, examples, demonstrations, etc.) to one of the lectures or topics. In fact, the success of modern LMS (Brusilovsky,Miller 2001) offering a simple interface for creating content folders and sorting the content into these folders, provide strong evidence that this approach is close to a teachers' own conceptualizations of course subjects.

An example of topic-based content modeling supported by the QuizGuide system is provided in Figure 1. Here, a course is structured into a sequence of *Topics* ordered along the course schedule. In simple cases, every *Topic* corresponds to a single lecture (or, in a more general sense, one of the top-level course *Goals*). However, it is not a rule – a teacher can allocate as much time for a topic as required to cover a coherent set of ideas and concepts encapsulated in a topic. As a result, one *Goal* (lecture) can combine one or more smaller *Topics*; however, if necessary, a complex but coherent *Topic* can be spread over several *Goals*. In total, QuizGuide's domain model consists of 22 *Topics* divided between 15 *Goals*. The educational content in QuizGuide is formed by a set of self-assessment *Quizzes* (their nature is explained in the next section). Following the topic-based content organization, each *Quiz* is associated with one of the course *Topics*. A *Quiz* is a composite educational activity – it consists of a sequence of individual traceable *Questions*. However, *Questions* can be accessed only sequentially within a *Quiz*; therefore, content assignment is done on the level of *Quizzes*.
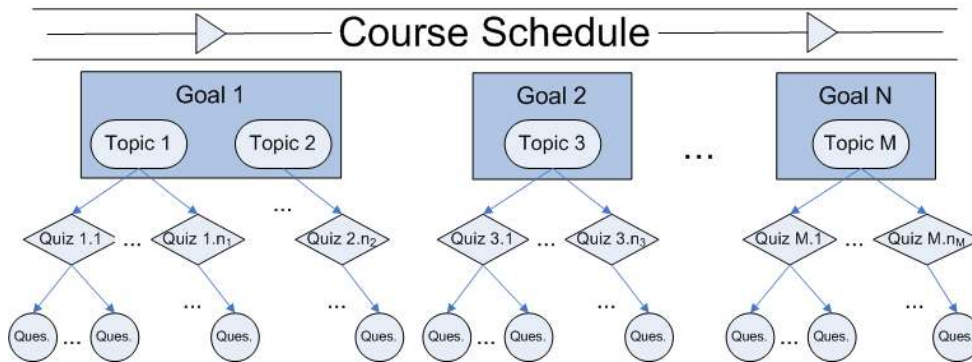
Figure 1: Topic-based content modeling in QuizGuide

While a topic-based content organization similar to the one shown in Figure 1 was used in numerous practical courses and textbooks, it was regarded as too simple for a meaningful personalization. Even those rare adaptive systems, which used topic-level domain models and grouped content into topics, tried to extend this basic approach in some way (Papanikolaou et al. 2003; Virvou,Moundridou 2001; Conejo et al. 2004; Stern,Woolf 1998; Eliot et al. 1997; Specht et al. 1997). Those extensions enabled more sophisticated personalization, but they also made the content models of the systems more difficult to manage. In contrast, the main challenge of our research was to explore the potential of the "native" version of the topic-based approach for UM and adaptation. Our challenge was twofold: (1) how to model student knowledge based on the coarse-grained domain structure and the history of student's interactions with topically organized educational content, and (2) how to use a topic-based model of student knowledge to provide meaningful and useful adaptation of educational content.

2.2.   TOPIC-BASED STUDENT MODELING

One goal of our project was to demonstrate that a pure topic-based course organization could provide a basis for meaningful student modeling on at least two levels – knowledge and goals. For modeling knowledge, we used a traditional weighted overlay student model (VanLehn 1988; Brusilovsky,Millán 2007), where student knowledge is represented as an overlay of the topic structure (i.e., for each topic, the system maintains an estimation of student knowledge of this topic). QuizGuide uses a numeric version of the overlay model (representing knowledge of each topic as a real number between 0 and 1) with no propagation between topics; however, more sophisticated fuzzy or Bayesian models could be used as well (Jameson 2001). The knowledge model is updated by tracking student activity with educational content. The update process is as simple and straightforward as the topic-based organization itself: each student's success or failure with a piece of educational content (a page is read, a question is answered, an animation is watched, etc.) causes a change in student knowledge level for the topic to which this content belongs.

To model learning goals, we also exploited an approach following everyday teaching practice: learning goals are mapped into the course lectures and organized along the course schedule (see Figure 1). In a typical course, each topic is presented in a specific lecture on a

specific date according to the course curriculum, and, at each moment of time, students are expected to focus on the current topic(s). To reflect this classroom practice, QuizGuide maintains the course schedule, which associates each topic with the date of its presentation. The ability to use a course schedule for goal modeling is a benefit of topic-level organization with its clear relations between topics and lectures. When annotating topics with adaptive icons, QuizGuide takes into account not only the individual progress of a student with the learning content, but also the association of topics with the current learning goal.

It is important to stress that the topic-based student modeling is universal. While the implementation of the approach presented in this paper has been done for a particular C programming course, it can be used with little or no modification in other domains and for other types of educational content. The focus of this approach is to track student knowledge and goals as an overlay of domain topics and to use the topic-based content organization to map student actions to student knowledge. Figure 2 visualizes the topic-based UM and adaptation approach as a general framework. Within this framework, we implemented a specific adaptation approach presented in the next subsection that uses adaptive link annotation as the key adaptation technology.
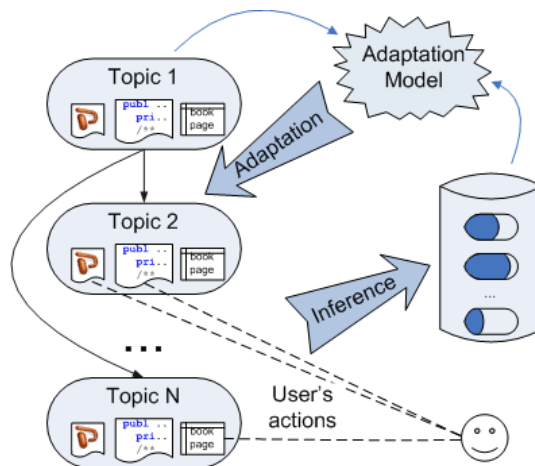

Figure 2. A topic-based adaptive system

2.3.   TOPIC-BASED ADAPTIVE NAVIGATION SUPPORT

The idea of topic-based adaptation in QuizGuide is directing students to the most appropriate topics by presenting the content of their individual student models in the form of navigational cues. Every topic link in QuizGuide is annotated with an icon combining the relation of the topic to the current learning goal and the current state of knowledge demonstrated with the corresponding learning material. As a result, a student is constantly aware of her performance and is able to focus on the parts of the course most important to her/him. This adaptation approach is combination of open student modeling (Bull 2004; Self 1990) and adaptive navigation support (Brusilovsky 2007).

Figure 3 presents a snapshot of the QuizGuide student interface. The left frame provides students with the annotated list of links to all topics available in the course. A click on a topic

link unfolds/folds links to the quizzes available for this topic. It allows students to organize their working space by opening only the quizzes that they want to focus on.
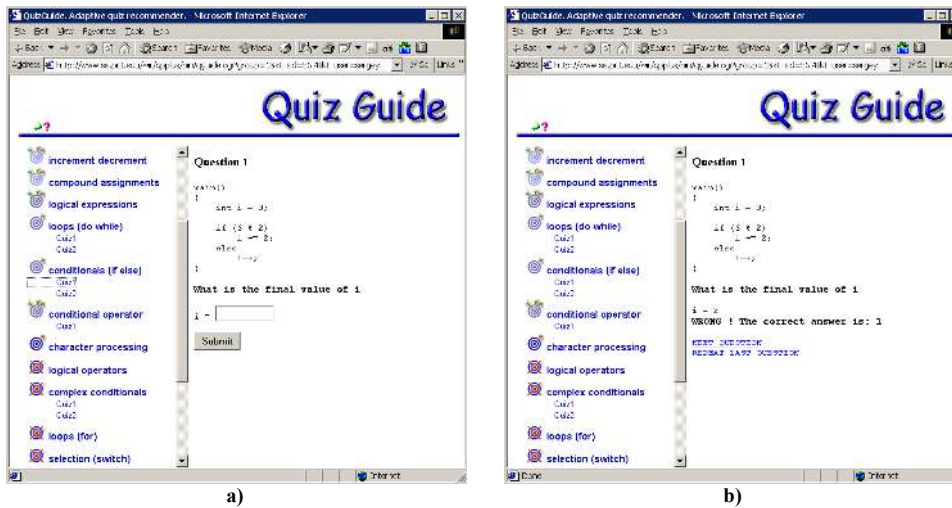

Figure 3. Interface of QuizGuide

The right frame of the interface contains a question that a student is currently working on. All quizzes and questions used in this study were provided by the QuizPACK system that was developed by our group (Brusilovsky,Sosnovsky 2005b). QuizPACK focuses on a special category of questions used in programming courses – code-execution questions. This type of question presents a student with a fragment of programming code and asks her/him to predict an outcome of this code (Figure 3a). QuizPACK can ask for two kinds of outcomes: the value of a particular variable or a string to be printed to the console by the code of the question. One of the core features of QuizPACK is the parameterized nature of its questions. One constant within a question code is randomly generated. As a result, every time a question loads, its content and expected correct response will be different. After the student gives an answer and hits the "submit" button, the system generates a feedback screen. This screen indicates the correctness of the answer (correct/incorrect) and provides the correct response, if the student has made a mistake (Figure 3b). It also presents the original text of the question; this allows the student to re-think the question and the answer. To attempt the question again with a different value of the generated parameter, the student can click the "REPEAT LAST QUESTION" link. To move to the next question, the student clicks the "NEXT QUESTION" link.

For annotating topic links, QuizGuide uses large "target-arrow" icons (Figure 4 summarizes them). The icons deliver two kinds of information to a student:
  – individual performance of the student on the quizzes associated with the topic
  – relevance of the topic to the current learning goal within the course
The number of arrows (from 0 to 3) in the target reflects the amount of knowledge demonstrated for the topic. It is a simplified view of the student knowledge model (a mapping of a positive real number representing a knowledge level into a simpler 0-3 integer scale). If no (or very little) knowledge has been demonstrated yet for the topic, the target will be empty, which

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Figure 3

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Figure 3

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Figure 4

invites the student to concentrate on this topic. Once the student starts giving correct answers to the questions of the topic's quizzes, the target-icon "starts receiving" arrows, indicating the increase in student knowledge. After a sufficient number of correct answers, the system collects enough evidence of student knowledge to annotate the topic with the "3-arrows target" – the highest level of mastery.

Figure 4. Topic-based annotations in QuizGuide: top – levels of topic goal, bottom – levels of topic knowledge

The *color* of the topic icon designates the relevance of the topic to the current learning goal. As the teacher of the course introduces new topics, QuizGuide annotates them with bright-blue icons representing the current learning goal of the course and inviting students to focus on them. Topics that have not yet been introduced are annotated with crossed-out target icons hinting that students might not be ready for their quizzes yet. Topics that have been introduced earlier in the course are no longer the focus of learning; QuizGuide indicates so by annotating them with two different pale colors. Topics that are prerequisites for any of the current learning goals are marked with pale-blue target icons. If a student has insufficient knowledge of those topics, she is likely to have problems with the current topics as well. Pale-blue annotations help students to remedy such situations by attracting their attention to the topics that are potentially useful within the context of the current learning goal. Pale-grey icons annotate past topics that are not prerequisites to the current topics. Thus, that the availability of topic prerequisites allows to distinguish between four goal states of a topic instead of regular three.

It is easy to observe that topic link annotation in QuizGuide combines two kinds of adaptation: individual progress-based adaptation and group goal-based adaptation. QuizGuide annotations inform students about individual and group-level value of the quizzes behind annotated links, trying to direct students to the best learning content at any particular moment of time. Nevertheless, following the adaptive hypermedia tradition, QuizGuide does not restrict access to educational content in any way. The students can access any topics, even those that have not been introduced yet.

QuizGuide employs icon-based (vs. text-based) link annotation, which helps students to immediately recognize a topic state, but also requires some time to remember the meaning of all annotations. To help students learn and understand annotations, QuizGuide dynamically generates mouse-over textual hints for all icons. The detailed help explaining all interface elements of the system is available as well.

3. Implementation of Topic-Based Student Modeling and Adaptation in QuizGuide

QuizGuide follows the traditional content-based adaptation procedure. It can be divided into three stages:
- The teacher defines the domain model (the topic-based structure of the course), the content model (connections between topics and quizzes) and the goal model (association of topics with the schedule of lectures/goals);

- UM server CUMULATE (Brusilovsky et al. 2005a) infers individual knowledge levels based on students' interaction with QuizPACK quizzes and the constructed content model;
- Adaptation module of QuizGuide merges individual models of student knowledge with the current goal of the course to generate adaptive annotations.

This section provides important details about the implementation of each of these stages in QuizGuide. It also outlines the architecture of the system to explain how different components interact with each other.

3.1.   ARCHITECTURE

Architecturally, QuizGuide is implemented as a value-added adaptive service. QuizGuide acts as an intermediary between the student and the learning content. Using the current state of each student's knowledge model, it generates adaptive link annotations. Potentially, it can work with any learning content, as long as the system proving the content is capable of reporting student actions to the user-modeling server. This paper focuses on QuizGuide application in C programming courses, for which the learning content (quizzes) is served by the QuizPACK system. However, QuizGuide has been also used to provide adaptive navigation support for SQL problems served by the SQL-KnoT system (Brusilovsky et al. 2010), and Java programming questions served by QuizJET (Hsiao et al. 2009). Figure 5 presents the architecture of QuizGuide and shows the flow of communication between the UM server (CUMULATE), the content provider (QuizPACK) and the adaptive service (QuizGuide). QuizPACK, working in a separate HTML frame, evaluates students' answers and sends records to CUMULATE using the standard HTTP protocol. Every record contains an exhaustive report about the attempt, including *user id*, *session id*, *application id*, *quiz id*, *question id* and the result of the attempt (correct/incorrect). CUMULATE augments this information with the timestamp and stores the record for model inference.

Peter Brusilovsky 7/27/2016 10:35 AM
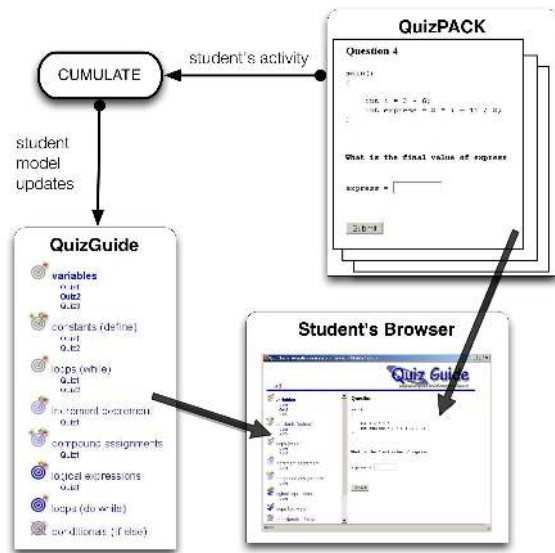**Deleted:** Figure 5

Figure 5. QuizGuide architecture

To generate annotations, QuizGuide sends an HTTP request to CUMULATE specifying the *user id* and the *domain id*. CUMULATE retrieves the relevant activity of the user, computes new levels of her knowledge for the topics and sends it back to QuizGuide as an XML report. QuizGuide extracts the knowledge levels from the XML report, merges them with the time-based learning goal information to identify current, passed, prerequisite and not-ready topics, and generates annotations. The annotated list is presented in the left frame of the student browser, while the question on which the student is currently working is loaded in the right frame.

3.2.   CONTENT MODELING

Every topic in QuizGuide is, essentially, a bag of content (quizzes). Each quiz is assigned to the most appropriate topic. The quizzes can have different influences on the topics to which they are assigned. Quizzes can differ in difficulty and coverage of the learning material important for mastering the topic. A teacher can model this contribution by assigning the percentage of the topical knowledge that a student can earn by completing a quiz. Once all quizzes are completed the topic is mastered. Consequently, the content model of QuizGuide contains the list of topics – each connected by weighted relations to the corresponding quizzes (Figure 6).

Peter Brusilovsky 7/27/2016 10:35 AM
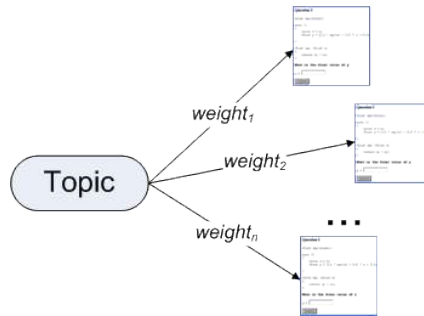**Deleted:** Figure 6

Figure 6. Topic-based content modeling in QuizGuide

The topic-based model of a course can either be a simple flat list ordered along the course schedule or have some structural relationships between the topics themselves. In order to take more advantage of the goal-based adaptation (i.e., generate four goal states), we have connected the C-programming topics to each other with the "*isPrerequisiteFor*"-relation. This relation models not the domain-driven structural connectivity, but the pedagogically-driven dependencies between pieces of associated content. In other words, if *Topic1* is connected to *Topic2* with the "*isPrerequisiteFor*" relation, it means that some of the questions inside *Topic2* require students to know learning material trained by some of the *Topic1*'s questions. For example, in the QuizGuide's model, the topic "*loops (for)*" is a prerequisite for the topic "*arrays*" (see Figure 7). It does not imply that in every possible implementation of the C-programming domain, arrays require knowledge of for-loop; but for this particular course, students working with the quizzes on arrays need to have knowledge of for-loop. Topic-based modeling speaks the language of a classroom teacher. It does not demand understanding the principles of domain semantics in order to connect topics to each other, but relies on the teacher ability to identify the pedagogical dependencies between topics used in the course.

Peter Brusilovsky 7/27/2016 10:35 AM
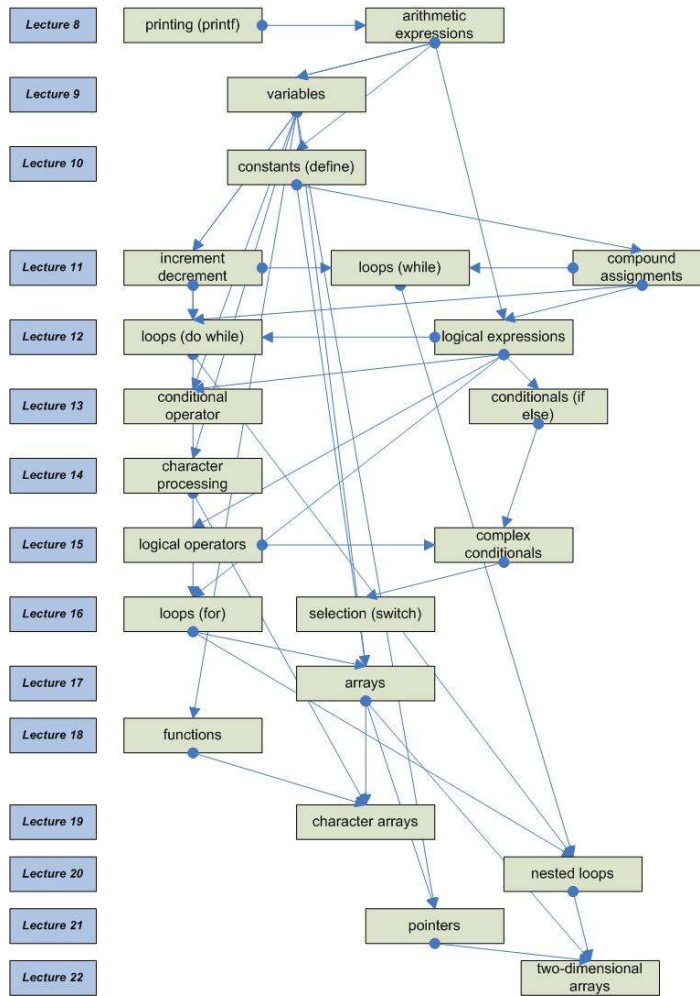**Deleted:** Figure 7

Figure 7. Topic-based structure of the C-Programming course (links represent *isPrerequisiteFor* relation)

3.3. USER MODELING

The maintenance and update of topic-based student models is performed by the user-modeling server CUMULATE (Brusilovsky et al. 2005a). CUMULATE represents information about a user on two levels: the event storage and the inferred model. The event storage accumulates information about all students interactions reported by content-providing applications (such as QuizPACK). The event storage is open to a variety of inference agents that process the event data in different ways and convert it into a more familiar form of name-value

pairs that altogether form the inferred user models. The architecture of CUMULATE allows developing multiple inference agents, which can compute different user parameters using different methods.

The inference agent used by QuizGuide implemented a relatively straightforward "*Average of sums of averages*" modeling formula for student knowledge computation:

$$K_i = \frac{\sum_{j=1}^{N_i} w_{ij} \frac{\sum_{k=1}^{M_j} x_{jk}/z_{jk}}{M_j}}{\sum_{j=1}^{N_i} w_{ij}}, \text{ where} \tag{1}$$

K_i – current level of knowledge for i^{th} topic,
N_i – number of quizzes participating in i^{th} topic,
M_j – number of questions in j^{th} quiz,
w_{ij} - weigh (contribution) of jth quiz in i^{th} topic
x_{jk} – number of correct attempts for the kth question of jth quiz,
z_{jk} – total number of attempts for the kth question of jth quiz.

This formula is consistent with the topic-based content modeling approach described above, as it measures the demonstrated knowledge on a topic by averaging the progress on its quizzes and questions. A student needs to complete all questions in a quiz to achieve the maximum knowledge level for this quiz. Similarly, if the student completes a quiz, her topic knowledge improves only by the fraction that this quiz holds in the entire topic. An important aspect of this formula is that the more events have been recorded for a topic, the smaller the impact of every new event on the estimation of the topic knowledge.

3.4. ANNOTATION ALGORITHM

The adaptive annotations of QuizGuide accumulate two kinds of modeling evidence: the individual knowledge level for a topic reflecting student knowledge on the associated quizzes and the relevance of the topic to the current learning goal shared by the entire course. Figure 8 presents a simplified algorithm computing an appropriate annotation for a single topic.

The knowledge level thresholds *mil_level*, *med_level* and *max_level* along with the weights can be assigned by the teacher. These thresholds allow fine-tuning the adaptation mechanism by defining how much knowledge a student needs to demonstrate before QuizGuide decides to add another arrow to the annotating target-icon. After analyzing the existing history of QuizPACK activity over the previous semesters and determining average patterns of successful and unsuccessful responses to its questions, we assigned rather lenient thresholds of 0.1, 0.3 and 0.5 to adjust for the sensitivity of the modeling formula to the negative evidence. The consequent evaluation of QuizGuide by students showed that, on average, they felt the annotations were timely.

```
If (topic.goal == current_goal)
       If (topic.knowledge < min_level) topic.annotation = ◎
       Else If (topic.knowledge < med_level) topic.annotation = ◎
       Else If (topic.knowledge < max_level) topic.annotation = ◎
       Else topic.annotation = ◎
Else If (topic.goal == prerequisite_goal)
       If (topic.knowledge < min_level) topic.annotation = ◎
       Else If (topic.knowledge < med_level) topic.annotation = ◎
       Else If (topic.knowledge < max_level) topic.annotation = ◎
       Else topic.annotation = ◎
Else If (topic.goal == past_goal)
       If (topic.knowledge < min_level) topic.annotation = ◎
       Else If (topic.knowledge < med_level) topic.annotation = ◎
       Else If (topic.knowledge < max_level) topic.annotation = ◎
       Else topic.annotation = ◎
Else   topic.annotation = ◎
```

Figure 8. Annotation algorithm in QuizGuide

4. Evaluating the Topic-Based Adaptation and User Modeling

While topic-based UM and adaptation approach forms the centerpiece of this work, the main focus of the paper is not on the approach itself, but on the evaluation of the approach. Trying to reach a compromise between the complexity of content organization and UM from one side and teachers' ability to structure the learning content and employ adaptation technologies, we considerably simplified the commonly used concept-based approach to adaptation and student modeling. We made both the domain model and the content model more coarse-grained and suggested relatively simple, easy-to-understand UM and personalization algorithms. With all that, we were especially interested whether or not we went too far, ending with an approach, which is more transparent for teachers, but failing to support meaningful personalization of educational content. To answer this question, we performed an extensive evaluation of the approach and the QuizGuide system implementing it.

The remaining part of this section reviews the employed evaluation methodologies and provides details about the evaluation context and data collection. The following sections report findings obtained using each of the methodology.

4.1. THE EVALUATION APPROACHES

Altogether, we used four approaches in our evaluation process. Two of them – the added value analysis (known as "with and without") and the user feedback analysis are traditional for the field of UM (Chin 2001). The *added value* analysis attempts to assess the effect of adaptation *as a whole* by contrasting an adaptive system and a comparable non-adaptive system. The goal of this analysis is to register some objective differences between the systems, (i.e., increase in speed of learning, decrease in error rate, etc.), which can be attributed to the presence of adaptation. In contrast to the objective nature of the added value approach, the user feedback analysis allows to assess the subjective impact of adaptation and to understand the impact of different system features on users' opinion about the system. While the data obtained from user feedback analysis

are considered less reliable (Gena,Weibelzahl 2007), they can be important to understand the true impact of an adaptive system.

The remaining two approaches – layered evaluation and navigation pattern analysis – are used less frequently in the literature. The goal of both approaches is to perform an "internal" analysis of a specific system or a technology. The layered evaluation approach (Paramythis,Weibelzahl 2005; Karagiannidis,Sampson 2000) attempts to assess separately various components of the UM and adaptation process while navigation pattern analysis approach (Srivastava et al. 2000) attempts to discover some holistic patterns of user behavior.

4.2. THE EVALUATION CONTEXT AND DATA COLLECTION

The evaluation of QuizGuide has been performed in the context of an undergraduate course "*Introduction to Programming*" offered at the School of Information Sciences of the University of Pittsburgh. The main part of this course was dedicated to introductory C programming covering the basics of variables, data types, control structures, character processing, arrays, pointers, and functions. Overall, 22 topics were taught during the 15 last lectures of the course (see Figure 7). Each lecture was defined as a learning goal in QuizGuide. We developed 155 QuizPACK questions combined in 44 quizzes to cover the course material. The topics incorporated between one and three quizzes. The quizzes contained from three to five questions.

The evaluation consisted of several studies of QuizPACK and QuizGuide over the period of four consecutive semesters. Each semester featured a different group of undergraduate students working with the systems in the context of the same course. The instructor, the syllabus of the course, the set of topics, the course books, the in-class quizzes, the homework assignments, and other characteristics of the course did not change during this time. During all four semesters, the role of QuizPACK/QuizGuide systems in the course also stayed the same; they were used as supplementary learning tools. The systems were introduced to students in the beginning of a semester as useful means to self-assess programming knowledge and prepare for regular in-class quizzes. Throughout the course, students were encouraged to use them, however participation in using the systems was voluntary and did not influence the course grades.

In the first two semesters, the corresponding groups were using only non-adaptive quizzes served directly by the QuizPACK system through a simple course portal (Brusilovsky 2004). Figure 9 presents a typical collection of learning content available within a lecture folder inside the course portal. Among links to other learning resources, the folder contains three QuizPACK quizzes (one of the quizzes is launched in the separate browser window).

In the middle of the third semester, QuizGuide was introduced. After that, for the second part of Semester 3 and the entire duration of Semester 4, the students had an option to use questions with adaptive navigation support – through QuizGuide (Figure 3) – or without adaptive navigation support – through the course portal (Figure 9). The set of quizzes and questions, as well as the structure of topics and lectures (learning goals), were the same for adaptive and non-adaptive modes and did not change over the time of the experiment.

Overall, 114 students participated in the experiment: 45 in Semester 1, 28 in Semester 2, 27 in Semester 3, and 14 in Semester 4. In non-adaptive semesters 1&2, 60 out of 73 students (82.2%) have tried the system at least once. In adaptive semester this number is 37 out of 41 (90.2%). We exclude from further analysis those students who have not answered a single self-assessment question. Altogether, during these four semesters, students answered more than 20,200 QuizPACK questions and spent with the system around 800 hours.
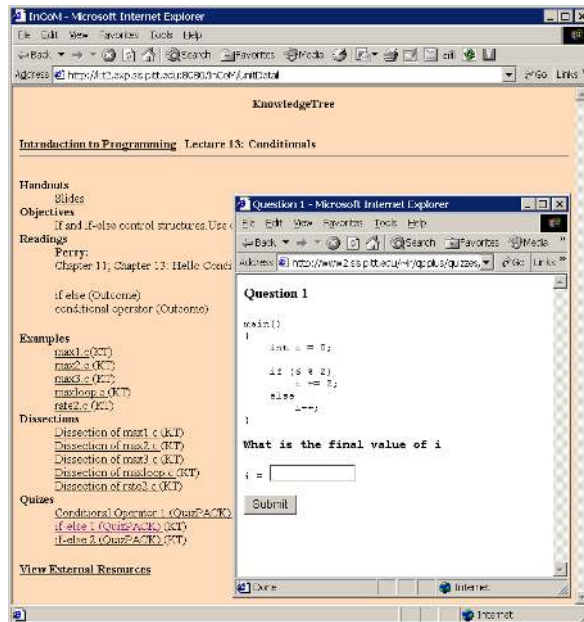
Figure 9. Non-adaptive access to quizzes

While most of the students taking the course never formally studied programming before, some students were capable of writing and launching simple programs prior to the course. To take into account students' starting level of domain knowledge, in the beginning of the course, they were asked to rank their own programming experience on the scale ("novice" – "medium" – "knowledgeable"). In addition, they took a pre-test consisting of 10 items specifically evaluating the subset of C programming knowledge taught during the course and covered by QuizPACK questions. To estimate students' involvement in the course, we collected their attendance throughout the semester. Table 1 summarizes the main descriptive parameters of the student groups.

To ensure that the effects observed during the evaluation cannot be attributed to the differences in student population across semesters, but indicate the added value of adaptive annotation, we analyzed the variance in distribution of these parameters. The results of the chi-square tests for independence showed that there was no significant difference between QuizPACK and QuizGuide groups in gender distribution $\chi^2(1, N = 97) = .43$, $p = .51$. The Mann-Whitney test[1] showed that the class attendances of students during non-adaptive semesters ($Mdn = 89.7\%$) and students during adaptive semesters ($Mdn = 81.8\%$) did not significantly differ from each other as well, $U = 1324$, $p = .11$, $r = .16$. For self-assessed programming experience, the Mann-Whitney test also did not find a significant difference between non-adaptive and adaptive groups, $U = 765$, $p = .26$, $r = .12$ (the mean ranks are 1.96 and 2.15, respectively). The

---

[1] In the reminder of this paper, Mann-Whitney test is used as a substitute of an independent-samples *t*-test, whenever our data do not satisfy the normality assumption.

difference between the pre-test score of the non-adaptive group ($Mdn = .5$) and the pre-test score of the adaptive group ($Mdn = 0$) is close to significance, $U = 1032$, $p = .054$, $r = .21$.

Overall, we can assume, that the two groups did not significantly differ one from another in terms of important demographic and background parameters.

Table 1. Student population parameters and their distribution across semesters.

		# of students	% of females	Attendance	Self-assessed experience (*1-3*)	Pre-test score (*0-10*)
Non-adaptive semesters	Semester 1	37	21.6%	$M = 84.3\%$	$M = 1.72$	$M = 1.55$
	Semester 2	23	21.7%	$M = 81.7\%$	$M = 2.26$	$M = .39$
	Total Non-adaptive	60	21. 7%	$M = 83.3\%$ $Mdn = 89.7\%$ $SD = 18.5\%$	$M = 1.96$ $Mdn = 2$ $SD = .77$	$M = 1.07$ $Mdn = .5$ $SD = 1.55$
Adaptive semesters	Semester 3	24	41. 7%	$M = 73.3\%$	$M = 2.13$	$M = .22$
	Semester 4	13	7.7%	$M = 88.3\%$	$M = 2.18$	$M = .75$
	Total Adaptive	37	29.7%	$M = 78.6\%$ $Mdn = 81.8\%$ $SD = 19.3\%$	$M = 2.15$ $Mdn = 2$ $SD = .66$	$M = 0.43$ $Mdn = 0$ $SD = .73$

5. The Added Value of Topic-based Adaptation: With-or-Without Evaluation

The most popular approach toward evaluating adaptive systems is "*with-or-without*" evaluation known also as *added value* or *adaptation effect* evaluation. Its main idea is to compare two versions of the target system: adaptive and non-adaptive. In the course of "*with-or-without*" evaluation, users interact separately with adaptive and non-adaptive versions working on comparable tasks in similar contexts. Various performance data for both versions are collected and contrasted. The difference in the results of these interactions is attributed to the effect (or *added value*) of adaptation.

A wide range of performance parameters has been used in the past studies. Most popular are directly recordable objective measure such as *time on task*, *error rate*, *knowledge gain*, etc. The choice of a performance parameter depends on the domain of the system, the user tasks and goals, and the nature of adaptation. For example, educational systems typically focus on knowledge gain, information access systems examine precision, and recommender systems focus on prediction error. To discover less noticeable effects of adaptation, such as time spent with an interface component or a number of particular actions (e.g. help requests), some studies use log analysis and data mining techniques.

To assess the impact of topic-based adaptation, we attempted to use both objective measures and log analysis. The educational nature of QuizGuide mandated to use *knowledge gain* as the primary measure of the added value of implemented adaptation on student learning. To estimate the effect of adaptive navigational cues on performance within system we employed the average success rate of student responses. Finally, we analyzed several parameters characterizing the amount of work done and time spent with self-assessment quizzes within a typical session and an entire semester. Given the non-mandatory usage of QuizGuide/QuizPACK, these parameters allowed us to assess students' motivation to use the system. Altogether, these measures helped us to confirm two different effects of topic-based

adaptive link annotation: the learning effect and the motivational effect, and, at least partially confirm its effect on within-system performance.

5.1. LEARNING EFFECT

For evaluating the impact of adaptive annotation on students' learning, we calculated their knowledge gains over the period of using the system. In order to do so, we administered pre- and post-quizzes. These quizzes consisted of 10 code evaluation questions (similar to those used in the system) covering most of the course material. The pre-quiz and the post-quiz questions were identical except for numeric values, which ensured that the correct answers were different. The pre-quizzes were taken by the students in the very beginning of the semester to estimate their initial level of expertise in the subject. The post-quizzes were administered after the last lecture to measure the increase in knowledge over the semester. The format and the content of the quizzes stayed generally the same over all five semesters.

The knowledge gain was calculated in two different ways. According to the canonical formula it is the difference between the corresponding results on a pre-quiz and a post-quiz:
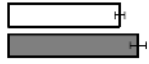
$$\textit{Knowledge Gain } = \textit{ PostScore} - \textit{PreScore} \qquad (2)$$

This formula sometimes is criticized for not adjusting to the differences in the students' initial knowledge levels. Therefore, another formula was introduced for calculating normalized knowledge gain (Hovland et al. 1949)

$$\textit{Normalized Knowledge Gain } = \frac{(PostScore - PreScore)/MaxScore}{1 - PreScore/MaxScore} \qquad (3)$$

As the evaluation showed, students working with "adaptive" version of the system achieved higher knowledge gain (see Table 2). There was a significant difference in knowledge gain for adaptive ($Mdn = 7$) and non-adaptive groups ($Mdn = 6$), $U = 539$, $p = .019$, $r = .26$. The effect also holds for the normalized knowledge gain: for the students working with QuizGuide ($Mdn = .7$) it is significantly higher than for the students working with QuizPACK ($Mdn = .6$), $U = 572$, $p = .044$, $r = .22$. This impact on learning can be easily explained when it is considered together with the motivational effect of the topic-based navigation support described in Section 5.3.

Table 2. Comparison of knowledge gain measures across adaptive and non-adaptive semesters

	Non-adaptive semesters	Adaptive semesters	Test statistics	Graphical representation
Knowledge Gain	$M = 5.77$ $Mdn = 6$ $SD = 2.06$	$M = 6.55$ $Mdn = 7$ $SD = 2.35$	$U = 539$ $p = .019$ $r = .26$	
Normalized Knowledge Gain	$M = .59$ $Mdn = .6$ $SD = .20$	$M = .68$ $Mdn = .7$ $SD = .23$	$U = 572$ $p = .044$ $r = .22$	

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Table 2

5.2. WITHIN-SYSTEM PERFORMANCE EFFECT

The added-value effect of adaptation can be observed not only outside the system, but also within it. By comparing the parameters characterizing students' performances with self-assessment quizzes accessed through QuizPACK and QuizGuide, we can measure whether, in the presence of adaptive navigation support, students are capable to self-regulate their learning more effectively and attempt quizzes that they are better prepared for. The most straightforward way to estimate the within-system performance is to compute the ratio of successful attempts on QuizPACK questions. Our hypothesis has been that QuizGuide helps students choose topics in a more meaningful way by (1) helping them to avoid topics they are not ready for (crossed-out icons), (2) inviting them to focus on topics that might be the source of their current learning difficulties ("prerequisite" annotations), and (3) raising the general awareness about their overall topic-by-topic knowledge and motivating them to achieve amore complete state of annotations (different numbers of arrows in the target icons). Thus, students working with QuizGuide should be able to achieve a higher success rate than those working with non-adaptive quizzes.

Table 3, presents the t-test results. Contrary to our expectations, there was no significant difference in the success rate between the adaptive ($M = .42$, $SD = .15$) and the non-adaptive ($M = .43$, $SD = .16$) semesters; $t(95) = .23$, $p = .82$, $d = .047$. We hypothesized that the expected performance differences might have been flattened by two factors. First, the groups working with the non-adaptive system had considerably stronger initial C programming knowledge (see last column of Table 1), which could have increased the within-system performance of the "non-adaptive" group. Second, since the students in the "adaptive" group had access to both QuizPACK and QuizGuide and the use of QuizGuide was not required, some students might have kept using the non-adaptive QuizPACK throughout the semester, thus flattening the expected performance increase.

Table 3. Comparison of success rate across adaptive and non-adaptive semesters

	Non-adaptive semesters	Adaptive semesters	Test statistics	Graphical representation
Success rate	$M = .43$ $Mdn = .43$ $SD = .16$	$M = .42$ $Mdn = .44$ $SD = .15$	$t(95) = .23$ $p = .82$ $d = .047$	

Indeed, when we took a closer look at the adaptive semesters only, we discovered that after QuizGuide was introduced, some students immediately switched to the new system (22%), others kept using the non-adaptive one (13%), while the rest (65%) continued using both the adaptive and the non-adaptive versions. To compare the success rates between more coherent groups of students, we limited the comparison to the students of the "adaptive" groups and divided them into two subgroups with respect to the system they used primarily for accessing self-assessment quizzes: those who answered more questions while working with QuizGuide and those who answered more questions while launching them from the non-adaptive course portal. The analysis of the success rates of these two subgroups (Table 4) shows that, during the adaptive semesters, the students who mostly accessed quizzes through QuizGuide on average have achieved significantly higher within-system success rate ($M = .47$, $SD = .13$) than the students who mostly accessed quizzes through the course portal ($M = .36$, $SD = .15$), $t(35) = 2.31$, $p = .027$, $d = .77$. Such a high value of Cohen's d – more than three quarters of the pooled

standard deviation – indicates a rather large effect size. This data provides good evidence that adaptive navigation support, indeed, helped students to choose topics more wisely and achieve higher performance with self-assessment quizzes.

Table 4. Comparison of success rate between students who mostly used adaptive version of the system and those who mostly used the non-adaptive one

	Used mostly non-adaptive	Used mostly adaptive	Test statistics	Graphical representation
Success rate	$M = .36$ $Mdn = .31$ $SD = .15$	$M = .47$ $Mdn = .51$ $SD = .13$	$t(35) = 2.31$ $p = .027$ $d = .77$	

5.3. MOTIVATIONAL EFFECT

The most dramatic effect of topic-based adaptation was related not to students' learning but to their motivation. The introduction of the open student modeling and adaptive navigation support caused an impressive increase in the amount of student interaction with educational content as compared to their work with the non-adaptive interface in previous semesters. Since their work continued to be non-mandatory, we refer to the observed phenomenon as the *motivational effect* of adaptive link annotation. We have observed this effect in several e-Learning systems that use progress-based adaptive link annotation (see (Brusilovsky et al. 2009) for the detailed analysis). In this section, we report a short summary of the motivational effect of QuizGuide.

With QuizGuide, students answered more questions, worked with questions more persistently and accessed a larger diversity of question material. In some sense, adaptive annotations made work with the system more "*addictive*": students came to the system more often and, once they started a session, they stayed with it longer.

To investigate this effect, we again compared the log data from the adaptive and the non-adaptive semesters, but, this time, we looked at the diversity of usage parameters characterizing student engagement and the amount of work. System logs recorded every attempt to answer a quiz question and gave clear evidence as to whether a student accessed a question through the adaptive service or not. Data collection procedures did not differ across the target four semesters and did not depend on the interface used to access quizzes (whether it was adaptive or non-adaptive). The user-modeling server processed student work with the both systems in the identical manner.

We used three variables to parameterize student performance:
1. Activity/Attempts: the total number of *answers* submitted to QuizPACK question,
2. Coverage/Quizzes: the number of *unique quizzes* taken, and
3. Coverage/Goals: the number of *learning goals* (lectures), for which a student attempted at least one quiz.

Each of these variables was aggregated on two levels:
- Semester performance level: the total number of attempts made, quizzes explored, and learning goals covered by each user during the semester; and
- Session performance level: the average number of attempts made, quizzes explored, and average number of learning goals explored per session.

In addition, we compared the number of sessions started by an average "adaptive" student versus a "non-adaptive" one.

The goal was to support our subjective observation that adaptive guidance does provide a motivation impact, i.e. to determine whether activity, quantity and coverage was higher for students who were exposed to QuizGuide than for those who used the non-adaptive QuizPACK interface. The results of our comparison demonstrate the value of the adaptive navigation support in motivating students to interact more with non-mandatory self-assessment learning content.

As can be seen from Table 5, all variables and aggregation levels showed a clear advantage of the adaptive semesters over the non-adaptive ones. On all the parameters, students exposed to the adaptive features of QuizGuide on average performed roughly 1.5–2 times better, and for most of them, Mann-Whitney tests demonstrated significant results.

The only two metrics, for which the level of significance was not reached, were the overall number of sessions and the semester-level goal coverage. On the former of these parameters, QuizGuide, actually, performed more than 60% better on average; however, due to very large variances in session length within both groups, p is rather high and r is low. The second non-significant effect can be explained by the course-level ceiling of the corresponding parameter: the total number of learning goals (lectures) in the course is 15. Yet, if we look at the quiz-based coverage that is much less affected by the ceiling effect (number of quizzes in the course is 44), students accessed almost twice more unique quizzes during adaptive semesters versus non-adaptive ones. This essentially means, that adaptation caused them to explore more quiz content within every lecture.

The difference in the raw activity demonstrates that students have invested significantly more efforts in the non-mandatory self-assessment with adaptive navigation present. And the per-session parameters show that an average adaptive session was significantly more intense, both in terms of the number of attempts and in terms of the diversity of accessed quiz content.

Table 5. Comparison of average system usage parameters across adaptive and non-adaptive semesters

		Non-adaptive semesters	Adaptive semesters	Test statistics	Graphical representation
Per semester	Activity/ Attempts	$M = 132$ $Mdn = 88.0$ $SD = 119$	$M = 279$ $Mdn = 126$ $SD = 313$	$U = 769$ $p = .011$ $r = .26$	
	Coverage/ Quizzes	$M = 13.7$ $Mdn = 14.0$ $SD = 7.87$	$M = 24.5$ $Mdn = 23.0$ $SD = 12.0$	$U = 544$ $p < .001$ $r = .42$	
	Coverage/ Goals	$M = 8.77$ $Mdn = 9.50$ $SD = 4.65$	$M = 10.5$ $Mdn = 11.0$ $SD = 3.62$	$U = 874$ $p = .079$ $r = .18$	
Per session	Activity/ Attempts	$M = 12.8$ $Mdn = 10.7$ $SD = 10.9$	$M = 18.7$ $Mdn = 16.3$ $SD = 11.2$	$U = 632$ $p < .001$ $r = .36$	
	Coverage/ Quizzes	$M = 1.89$ $Mdn = 1.71$ $SD = .73$	$M = 3.71$ $Mdn = 2.79$ $SD = 2.76$	$U = 253$ $p < .001$ $r = .65$	

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Table 5

	M = 1.40	*M* = 2.08	*U* = 533	
Coverage/ Goals	*Mdn* = 1.32	*Mdn* = 1.67	*p* < .001	
	SD = .40	*SD* = 1.53	*r* = .44	
	M = 11.2	*M* = 18.0	*U* = 939	
Number of Sessions	*Mdn* = 8.50	*Mdn* = 11.0	*p* = .20	
	SD = 9.21	*SD* = 28.3	*r* = .13	

6. A Layered Evaluation of the Topic-Based Approach

In the previous sections, we followed the "with or without" evaluation approach that has long been considered the golden standard in the field of adaptive systems (Chin 2001). We remind that the idea of this approach is to engage two groups of subjects into using two different versions of the evaluated system – with and without adaptation. Then the comparison of the usage and performance parameters between the groups is considered as the key evidence about the effectiveness of the implemented adaptation technology. However, what is actually evaluated in a "with or without" study of QuizGuide driven by a topic-based UM approach? Is it the specific way to provide adaptive navigation support using target-arrows metaphor, the topic-based approach toward modeling student knowledge, or just the quality of the job done by the author in defining topics and connecting them with quizzes and questions?

To provide a deeper analysis of the topic-based approach exploited by QuizGuide, we decided to apply a layered evaluation approach (Brusilovsky et al. 2004; Paramythis,Weibelzahl 2005), which advocates the need to evaluate separately different aspects and components of the personalization process that contribute to its success. Main goal of our layered evaluation was to assess separately the UM (Sections 6.1 and 6.2) and the adaptation (Section 6.3) parts of an adaptive system. In addition, we set apart the evaluation of the underlying *domain model* (Section 6.1) from the evaluation of the *student-modeling process* itself (Section 6.2). This is important separation for evaluation of AESs employing *overlay student modeling* (such as Intelligent Tutoring Systems or Adaptive Hypermedia), since the quality of overlay student models to a considerable extent depends on the quality of an underlying domain model. Following this need, we start with the first evaluation stage – by assessing topics as modeling units, i.e., their applicability to serve as elementary pieces of domain semantics for building overlay models. After that we move to the second stage of the user-modeling evaluation that measures – how meaningful the resulting models are, how precisely they mirror the actual state of students' knowledge. To measure the quality of domain modeling, we use learning curve approach. To measure the quality of UM, we assess the average predictive validity of populated student models. As mentioned above, the second stage is not completely independent from the first one. If topics are mediocre assessment units the predictive validity of topic-based user models is inescapably low.

6.1. EVALUATION OF TOPIC-BASED USER MODELING: TOPIC AS AN ASSESSMENT UNIT

To evaluate how well the topic-based inference agent of the CUMULATE user-modeling server assesses student knowledge, we first need to investigate whether the selected topics are acceptable units of knowledge assessment. The chosen modeling approach, as well as the implemented adaptation strategy, could be reasonable; however, the resulting adaptive behavior might not be adequate to the student's actions and expectations if the assessment units are wrong.

For the evaluation of large topics as knowledge assessment units, we have applied learning curve analysis (Martin et al. 2011). Multiple experiments provide strong evidence that on the cognitive level, learning process generally follows the power law (Anderson et al. 1995; Newell,Rosenbloom 1981). In other words, the error rate of a learning skill decreases as the power function of the number of learning steps involving this skill[2]. The assumption behind the evaluation approach based on learning curve analysis is that learning curves that better approximate power law correspond to more cognitively valid units of knowledge. This approach has been pioneered in several projects (Koedinger,Mathan 2004) and formally suggested as an evaluation approach in (Martin et al. 2005). Since that, it has been used in a number of publications to assess the quality of domain modeling. A good summary of this approach could be found in (Martin et al. 2011).

To compute topic-based error rates, we have categorized all question attempts per student, per topic. Hence, if a student gave 20 answers to question on *Arithmetic Expressions*, 20 entries have been created in the order the answers were given; the actual questions did not matter (it could be 20 answers to the same question or to 15 different question on the topic of *Arithmetic Expressions*. Table 6 presents an extract of these data.

Table 6. Data for computing topic-based error rates

StudentId	Topic	Step#1	Step#2	…
24	*Printing (printf)*	*0*	*1*	*…*
24	*Arithmetic expressions*	*1*	*1*	*…*
…	*…*	*…*	*…*	*…*
26	*Printing (printf)*	*0*	*0*	*…*
…	*…*	*…*	*…*	*…*
Average error rate		*0.59*	*0.54*	*…*

Correct answers are marked with *1*'s, incorrect – with *0*'s. Data from all semesters (both adaptive and non-adaptive) have been used. The maximum number of learning steps per student per topic is 214; the average is 11.86. The following formula has been used to compute *Average error rates* for every *Learning step*:

$$AverageErrorRate = \frac{NumberOfIncorrectAttempts}{TotalNumberOfAttempts} \qquad (4)$$

Figure 10 demonstrates the dependency between the *Average topic-based error rate* and the number of *Learning steps*. Formula 5 defines the standard power-law learning curve approximating this dependency. The downward trend shows some learning effect and R^2 statistics indicates that about 39% of the variability in the *Average topic-based error rate* could be explained as the power law on the number of *Learning steps*.

$$TopicBasedErrorRate = .64 * LearningStep^{-.07}; \qquad (R^2 = .39) \qquad (5)$$

---

[2] Some researchers assert that the learning curve should be approximated with the exponential instead of power function (e.g. (Heathcote et al. 2000; Murre,Chessa 2011))

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Table 6

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Figure 10

This fit, however, is lower than results reported in the ITS literature that typically deals with smaller units of knowledge. For example, Mitrovic and Ohlsson (1999) applied learning curve analysis to evaluate the quality of a very fine-grained constraint-based domain model of SQL-Tutor and obtained $R^2 = .73$ when fitting a power law curve for the data of 100 random constraints. Koedinger and Mathan (2004) evaluated several skill-level models in the context of an Excel spreadsheet tutor with learning curves and obtained results ranging from $R^2 = .59$ to $R^2 = .96$.
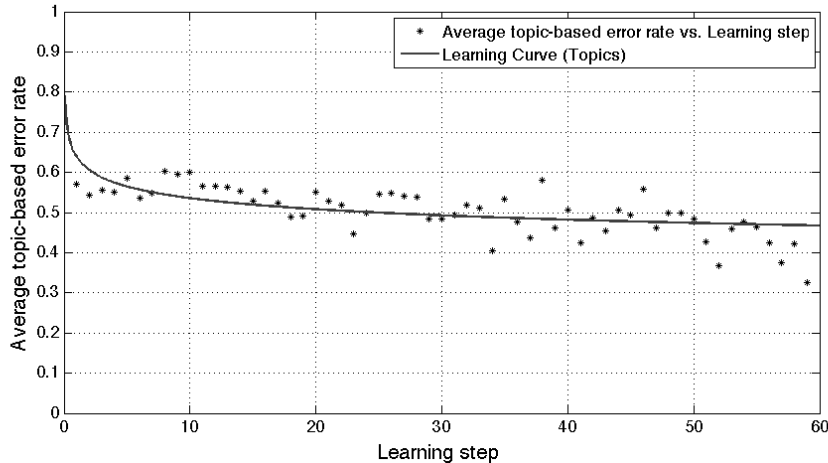

Figure 10. Topic-based learning curve.

As the analysis shows, the main feature of topics that differentiate them from the finer-grained concepts constraints and skills is that topics are large; they aggregate too diverse learning content and cover too much knowledge. To see, how the granularity influences the quality of topics as assessment units, we have performed the same learning-curve analysis on the smaller chunks of learning material, namely quizzes and individual questions (every QuizGuide topic combines on average 2 quizzes and approximately 7.5 questions). Figure 12 demonstrates that by narrowing the scope of assessment units we do achieve better results. The dependency between the *Error rate* and the number of *Learning steps* for an average quiz (Figure 12a) and an average question (Figure 12a) demonstrate better learning. Approximation of these plots with power law learning curves (Formulas 6 and 7) results in better fits (48% of the variability in the quiz *Error rate* and 70% of variability in the question *Error rate* can be explained using the power dependency on the number of *Learning steps*). These data demonstrate that QuizGuide topics, while being meaningful units of knowledge assessment, still cover too much learning content to reflect the changes in student knowledge as reliable as finer-grained units can do.

$$QuizBasedErrorRate = .66 * LearningStep^{-.12}; \qquad (R^2 = .48) \qquad (6)$$

$$QuestionBasedErrorRate = .64 * LearningStep^{-.21}; \quad (R^2 = .70) \qquad (7)$$
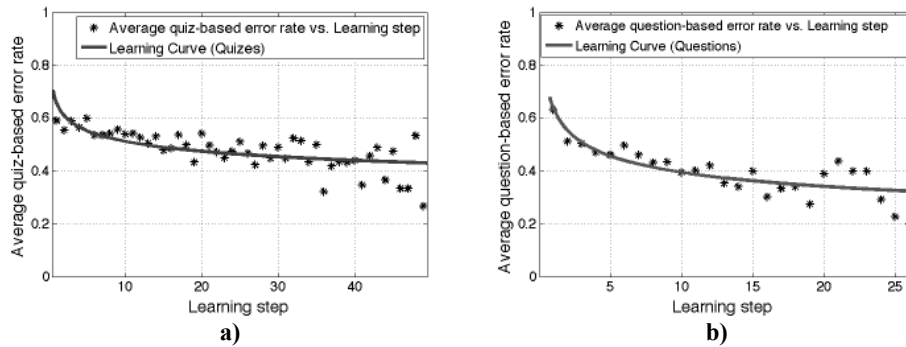
Figure 11. a) Quiz-based learning curve; b) Question-based learning curve

Since large topics involve too much knowledge, the precision of the assessment performed by the system might be reduced. When taking a quiz on a particular topic, a student actually practices a set of closely related concepts covered by the questions in the quiz. As a result, knowledge modeling on the level of a large topic is not capable of reflecting subtle changes in student's topic knowledge resulting from this practice. The lower quality of topic-based student modeling is an inherent feature of topic-based approach, however understanding its nature, i.e., why and how this quality is reduced, could help the designers of topic-based systems to minimize negative impact of large knowledge units. In particular, as demonstrated by the analysis above, a good way to increasing the quality of topic-based modeling is to keep topics reasonably small.

Another approach we have tried for improving topic-based knowledge modeling is to increase the predictability of students' results by harmonizing the assessment material. One of important aspects of learning content representation that our simplistic topic-based model missed on is the difference among questions in difficulty and structural complexity. A topic covering a dozen of questions treats all of them identically, even though students are much more likely to answer easy and simple questions correctly. By reducing of the variability in question difficulty within a topic one could improve the predictability of students' answers and hence improve the assessment characteristics of a topic. From another point of view when the question is too hard or too simple, the effectiveness of both learning and knowledge assessment is reduced.

To investigate this problem, we estimated the average question difficulty using the traditional measure – the mean value of error rate. The difficulty value ranged from .06 to .89, with $M = .46$ and $Mdn = .47$. We explored our data on two intervals traditionally used for estimating main trends that could be explained by the central part of the distribution: 90% and 50%. Figure 12a shows the very same learning curve as Figure 10, where 5% of the most difficult and 5% of the least difficult questions are removed from the computation of the average error rate. Figure 12b demonstrates a similar plot for the situation when all questions below the 25th and above the 75th difficulty have been filtered out. Formulas 8 and 9 define the equations for the power law learning curves approximating these plots. The error rate variability explained by the power dependence on the number of learning steps is almost double compared to the non-filtered topic-based learning curve (Formula 5).

$$TopicBasedErrorRate\_Filtered_{90} = .65 * LearningStep^{-.08}; \quad (R^2 = .60) \tag{8}$$

$$TopicBasedErrorRate\_Filtered_{50} = .69 * LearningStep^{-.13}; \quad (R^2 = .67) \tag{9}$$
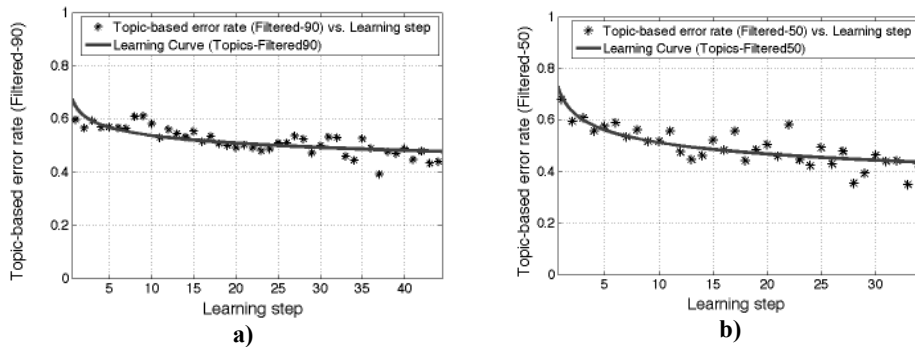

Figure 12. Learning curves for the average topic with 10% (a) and 50% (b) of "bad" questions filtered out.

There are several possible outcomes of this analysis that can be explored in order to improve topics' knowledge modeling quality. A more careful design of questions would lead to eliminating assessment items with "extreme" difficulty altogether. On the other hand, a more accurate manipulation of question difficulty can help decreasing the number of attempts on too difficult and/or too simple questions.

In summary, our analysis demonstrated that QuizGuide topics are meaningful units of knowledge assessment, but the larger size of topic make makes it hard for topic-based model to track student knowledge as reliable as it is done by finer grained knowledge units such as rules or constraints. In the following sections we investigate whether a reduced precision of topic-based modeling might become an obstacle for student modeling and personalization. However, our analysis also uncovered the mechanism of quality degradation and demonstrated that understanding this mechanism can lead to better topic and content engineering, which, in turn, could increase the quality of modeling.

6.2.   EVALUATION OF TOPIC-BASED USER MODELING: PREDICTIVE VALIDITY

The ultimate task of a student model is to represent the state of a student as close to reality as possible in order to help the AES anticipate student's actions and optimize her learning. If the predictions of a student model do not correlate with what students demonstrate in reality, then the quality and value of such a model is rather questionable. As a result, evaluating the *predictive validity* of a student model, i.e., its ability to predict the observed student performance based on the current model state, has become an important approach to measure the quality of student modeling in AESs. The predictive validity approach has been originally suggested by Corbett et al. (Corbett et al. 1993b; Corbett et al. 1993a) in the context of their work on student modeling approach known as on Bayesian Knowledge Tracing (BKT) for the domain of Lisp programming. Within years it has been adopted to evaluate other student modeling approaches, for example, SMART (Shute 1995), and explored in several other domains from algebra to genetics. Over time, predictive validity emerged as the standard approach to evaluate and

compare student modeling approaches. An important property of this approach in the context of our layered evaluation is its ability to isolate the quality of modeling from the effectiveness of adaptation.

In their original work on predictive evaluation of BKT, Corbett and Anderson distinguished internal and external predictive validity (Corbett,Anderson 1995). External validity is the ability to predict the results of student performance outside of the system (for example, post-test performance) while internal validity is the ability to predict accuracy of student performance within the system. In a number of early papers external validity was assessed in parallel with internal validity or even alone (Shute 1995), however, recent research almost exclusively uses internal validity assessment since it usually relies on much larger volumes of data and allows to assess student models on a finer grain level. It has become common to use internal validity not only to assess the quality of a specific approach in isolation, but also to compare several versions of the approach or compare a novel approach to a baseline. Following this trend, we performed an internal predictive validity evaluation of the modeling formula used by CUMULATE/QuizGuide, i.e., investigated how well topic-based knowledge levels produced by this formula predict the consequent performance of students.

As the baseline in our evaluation, we used BKT, the very approach where the predictive validity evaluation has been originally used (Corbett,Anderson 1995). BKT is one of the most popular techniques for modeling student knowledge in terms of elementary domain components. It predicts the probability of a student knowing a concept step-by-step based on the prior probability, the last result of the concept application by this student and probabilities of guessing and making a careless mistake. The original BKT and its many variations are widely used in research on Intelligent Tutoring Systems (ITS) and Educational Data Mining (EDM) making BKT the most popular baseline. For example, to evaluate a number of extensions and refinements of the original BKT approach, the authors usually compare the results of internal predictive validity evaluation of a novel technique against the original BKT on the same set of student logs (Beck et al. 2008; Baker et al. 2008; Pardos,Heffernan 2011; González-Brenes et al. 2014) (Pavlik et al. 2009; Yudelson et al. 2013).

To perform comparative predictive validity of QuizGuide student modeling, the states of QuizGuide student models as well as the BKT topic-based estimates were computed at each of all 18128 learning steps from the complete logs of our four semesters. For all learning steps of every student, we compute:

 – the prior knowledge level for a topic based on the QuizGuide formula;
 – the prior knowledge level for a topic based on the BKT formula;
 – the actual posterior student performance for the topic.

The prior knowledge levels for *QuizGuide* student models were computed using the original formula explained in Section 3.3 (i.e., in exactly the same way as it was done in all classroom studies). The prior knowledge for *BKT* student models was computed with the help of the toolkit described in (Yudelson et al. 2013) with by-skill parameter fitting using the gradient descent method.

To estimate the *actual performance* demonstrated by the student onward, we need to take into account the results of her attempts starting with the current learning step. However, correlating knowledge predictions with unprocessed single step performances is problematic. The size of topics, and differences in individual question difficulty (ranging from .06 to .89; see Section 6.1), contribute to a large variability in terms of student answers correctness. At the same time, QuizPACK questions do not account for partially correct student answers, which could

have helped smooth-out the short-term fluctuations in performance); therefore, on every step, a student response can be either correct (performance = 1) or incorrect (performance = 0). To address this problem, we have computed student posterior performance using the *simple moving averages* of student individual scores on the interval $n = 5$.

To see how well the models can predict/follow changes in student knowledge, first, we take a look at the correlation between the modeled values and the actual student performance. The results of the Pearson correlation analysis show that the ability of the topic-based knowledge model used by QuizGuide to predict immediate student's performance is fairly low ($r = .23$). The BKT formula produces a slightly better result ($r = .38$).

The reason for a weak correlation becomes clear, once we compare the histograms of all three measures (see Figure 13). While the actual performance values are almost uniformly distributed between 0 and 1 (middle histogram), the distribution of QuizGuide modeling values (left histogram) has the exponential shape with the majority of them staying between 0 and .25. At the same time, the distribution of BKT predictions (right histogram) is almost mirrored compared to the original QuizGuide formula. Most of its values lie in the interval from .6 to .75 with a long tail toward zero.

This picture can be further explained by analyzing the two modeling formulas. The formula used in QuizGuide (see Formula 1) has been designed from a teaching perspective, not a knowledge engineering one. Teachers tend not to trust student prior knowledge and need some solid assurance that a student *really* knows a topic if it is reported as known. The QuizGuide knowledge modeling starts with zero and favors students trying to solve all the questions in the topic correctly, even though that might not always be necessary to master the topic. The two main drawbacks of this formula are: (1) it is conservative: student might learn faster than the their modeled knowledge level grows; (2) it is not forgiving: if a student has accumulated a history of unsuccessful attempts while solving some topic's questions, it will become hard for her later to "prove" her mastery to the formula, as it treats all events equally without decaying the importance of the older evidence. Due to these aspects, the original QuizGuide formula tends to underestimate student knowledge staying on the "safe side", just as many teachers do[3].

In its turn, the BKT formula has been originally aimed at correct estimation of the probability that a student has mastered a skill as a result of practicing it. Unlike QuizGuide modeling approach, the initial BKT knowledge levels fitted to our data range from .003 (topic Arrays) to .94 (topic Compound Assignments) with M = .37 (SD = .23). It is much more reactive to changes in student knowledge: once a student demonstrates a sequence of correct attempts for a topic, the BKT prediction will quickly exceed the .5-threshold. However, due to the size of topics and variability in question difficulty, BKT tends to overestimate student ability to answer a new question correctly, especially, if the student has demonstrated a series of successes on previous (possibly easier) questions of this topic. Note that the tendency of BKT to overestimate student knowledge has been observed in other domains as well (Corbett et al. 2000).

---

[3] The reviewed drawbacks of the original topic-based student modeling formula caused us to replace it with a more cognitively grounded formula in the newer version of topic-based adaptation service (Yudelson 2010), however, in this paper, we analyze the exact formula that was used in our QuizGuide studies.
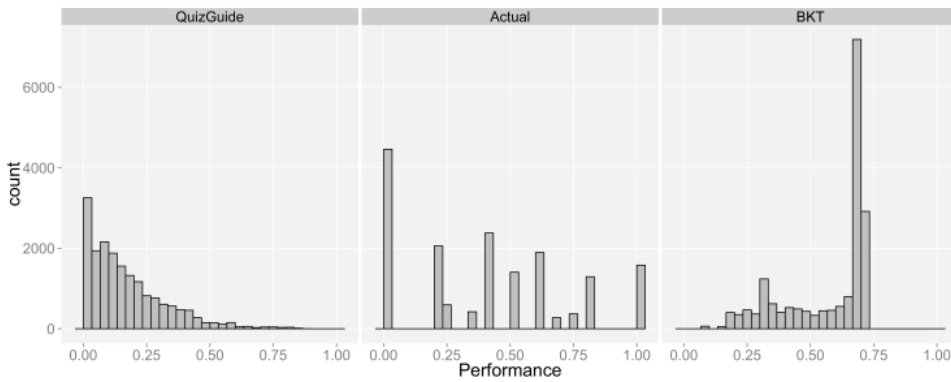
Figure 13. Histograms of (1) topic-based estimations of student knowledge using QuizGuide formula, (2) topic-based immediate actual performance of students averaged over five next attempts, and (3) topic-based estimations of student knowledge using BKT formula.

Further comparison of the two modeling formulas was done using *ROC curve* – a popular tool from the signal detection theory that recently became widely used in HCI research for evaluating reliability of data-driven models. An ROC curve graphically represents the quality of a binary classification model by plotting its true and false positive prediction rates for different values of the operating threshold. As the threshold decreases from 1 to 0, the model can recognize more and more positive cases correctly, but it also starts classifying more and more negative cases as positive. More information on the ROC curve analysis can be found e.g. in (Fogarty et al. 2005).

Figure 14 presents ROC curves of both the QuizGuide model and the BKT model on the same plot. It is easy to notice that BKT outperforms the original QuizGuide model. Its ROC curve is higher, meaning that for the same false positive rate it has a higher true positive rate. In other words, for the same percentage of cases where BKT *incorrectly* classifies a topic as mastered (predicts that a student will answer a question correctly), as it has a higher chance to *correctly* predict student mastery than the QuizGuide formula. Nevertheless, both curves are above the *TPR = FPR* diagonal, which means both models are better than random chance predictions. To numerically estimate the quality of the models and the difference in their quality, we have computed Area Under ROC (AUROC) values and their standard errors. This allows us to compute a *z*-test for means. Table 7 summarizes the results. Both QuizGuide and BKT models produce significantly better predictions than we would obtain by guessing. At the same time, BKT model significantly outperforms the original QuizGuide formula in terms of predictive validity.
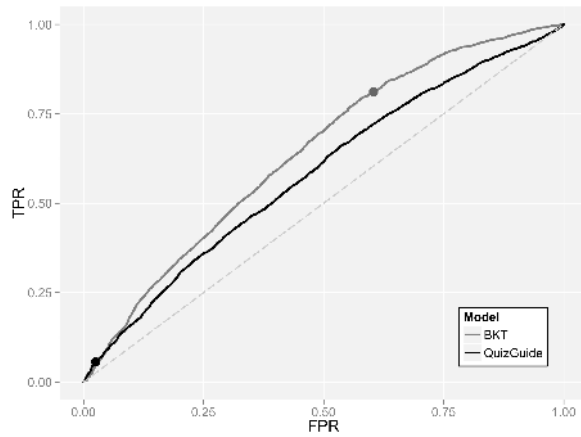
Figure 14. ROC curves for QuizGuide and BKT models with the .5 operating thresholds highlighted

Table 7. Area under ROC results for QuizGuide and BKT formulas

	AUROC	Compare to Random Chance	Compare to Each Other	Graphical representation
QuizGuide	$M = .58$ $SE = .004$	$z = 19.7$ $p < .001$		Threshold = .5
			$z = 9.8$ $p < .001$	QuizGuide
BKT	$M = .64$ $SE = .004$	$z = 34.5$ $p < .001$		BKT

The dots on the ROC curves mark the .5 operating thresholds of the both models. The upper parts of the curves (above the .5 point) represent the data obtained with lower thresholds – when the models are more likely to classify a topic as mastered both correctly (TPR) and incorrectly (FPR). Consequently, the lower parts correspond to the thresholds increasing from .5 to 1, while models produce fewer and fewer positive predictions. It is easy to see that as the operating threshold approaches .5, QuizGuide model rarely classifies a topic-user pair as a case of "perfect" mastery. This provides another perspective on the problem visualized by Figure 13 (left plot) and once again confirms the conservative nature of the original QuizGuide modeling formula. It is important to note, however, that when the student modeling knowledge estimates are converted into the navigational cues, QuizGuide effectively compensates for its conservative modeling approach by applying an "encouraging" scale of adaptation where .5 knowledge level corresponds to the maximum-knowledge icon[4]. As a result, the combination of modeling-adaptation performs quite meaningfully and responsively – while the modeling itself is conservative, visualized knowledge levels change faster than the modeled growth of knowledge.

---

[4] Figure 8 explains why .5 is the maximum level of student per-topic knowledge that results in a change in system adaptation: at this level the topic is annotated with the 3-arrows knowledge icon; any additional (positive) results will not influence the change of the topic annotation.

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Figure 13

One arrow could be obtained with just a few successful results and three arrows (unlike "perfect" mastery) are well reachable for students who want to master the topic.

The BKT formula, on the opposite, is, generally, more "optimistic" about student mastery: the part of the BKT's ROC curve above the 0.5 operating threshold is much shorter than its complement, which means. It is interesting, that the only place where the QuizGuide model outperforms the BKT model is the lowest segment where it is hard to make a positive prediction. In these conditions, the conservative QuizGuide model is a little more accurate. It is interesting that with its overall better precision, BKT approach will not work in QuizGuide unless the adaptation thresholds shown in Figure 8 are adjusted to the effective scale of BKT modeling. With the current thresholds all topics will be annotated with maximum level knowledge most of the time. This observation demonstrates that a more precise student modeling approach might not work better on practice than a less precise one, it is the combination of an approach itself and an adaptation strategy built on the top of it that makes the difference.

In this context, an interesting discussion point is whether the conservative nature of QuizGuide topic modeling is a weakness or strength in a system, which uses adaptive navigation support as the key adaptation approach. It is important to note that the original BKT approach was designed for *mastery learning* (Corbett,Anderson 1992) adaptation, which uses the achieved level of knowledge to determine when to stop practicing specific set of skills. In this context, the student is not involved into the decision process and high precision is valuable to stop unproductive work with sufficiently mastered skill. In contrast, QuizGuide's *adaptive navigation support* type of adaptation uses knowledge level to select adaptive link annotation, which, in turn helps students to select most appropriate topics. In other words, it is not the precision of the student model itself, but a student perception of the displayed knowledge level that makes the difference. In this context, a more conservative teacher-imitating knowledge estimation might provide a better motivation to work. We will further discuss this issue after examining the impact of the visual cues and student perception of modeling accuracy.

6.3.   EVALUATION OF TOPIC-BASED ADAPTATION: THE IMPACT OF VISUAL CUES

The final stage of layered evaluation is the evaluation of *adaptation* itself. It needs to be distinguished from the evaluation of the *adaptation effect* or the *added value of adaptation* (see Section 5). Instead of measuring the pragmatic utility of the adaptation, and the ways it helps students' learning, this section concentrates on analyzing the adaptation per se, addressing such issues as "*Does adaptation make the difference in how students interact with the system?*" and "*Do the students react on adaptive interventions as presumed?*" There is no single recipe for finding answers to these questions. It would differ not only for every adaptation technology, but also for different implementations of the same technology. For example, in the case of *adaptive search*, where the product of adaptation is an individualized ranking of search results, it is natural to check whether the resources with highest personal relevance placed on the top has a higher chance to be noticed, and clicked (Keane et al. 2008; Ahn et al. 2008). In the case of *social navigation support* where link annotation is used to indicate most socially valuable resources, it has been suggested to check whether the favorably annotated resources do receive more visits from the target users (Farzan,Brusilovsky 2008).

In the case of QuizGuide, an adaptive icon communicates two things to a student:
– how much knowledge progress she has accumulated for the relevant topic (this section focuses on analyzing how this information impacts student behavior);

–   how relevant the topic is to the current learning goal (the effect of this information is tested in Section 7).

QuizGuide's interface has been designed to guide the students to the topics where they currently underperform. Consequently, the intended behavior for a student following QuizGuide progress-based annotation could be described by the following rule: "*prefer topics annotated by icons with fewer arrows*".

In order to see how the progress-based adaptation of QuizGuide affects students' choices of quizzes to work with, we have had to take a closer look on what constitute these choices. At every moment of time, a student deciding which question to try next has 22 course topics to select from. The user model keeps track of student's history and can estimate her current knowledge for all the topics. A student accessing questions through QuizGuide is constantly informed of her demonstrated knowledge for all the topics by being presented with a combination of "0-3 arrows"-annotations. The "optimal" choice is to always prefer less mastered topics with a lower level of annotation[5] before moving to the next more advanced topics.

The easiest way to compute a metric showing how well a student follows the navigation would be by simply counting how many times topics with a certain number of arrows on the icon have been accessed. We used this approach in (Sosnovsky,Brusilovsky 2005) and showed that QuizGuide students, indeed, accessed more frequently topics with smaller number of errors. However, this metric does not account for the fact that at every moment the numbers of topics with different levels of annotations are very different. In the beginning, all topics are annotated with zero arrows, as a student has demonstrated no knowledge of them yet; while later in the semester, she really can make "*good*" (few arrows) and "*bad*" (many arrows) topic choices.
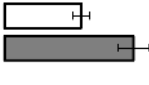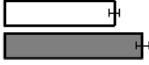
Hence, instead of concentrating on the fact that a topic with a particular visual cue has been chosen, we have had to concentrate on the decisions that students have made, taking into account not only the level of the chosen topics, but also the levels of all other topics in the course at the moment when this decision was made. We have re-ran the UM process on the logs of non-adaptive and adaptive courses, and, at every step, for every user, the snapshot of her knowledge state for all course topics have been restored and recorded. As a result, on every step, we could estimate the quality of a navigational decision a student has made. Two different metrics have been computed in order to quantify the average percentage of "good" navigational decisions made by every student.

- The *metric-A* is an aggressive metric. It assumes that students should always make the best decision possible. Only if a student choses a topic with the lowest level of annotations (minimum number of arrows), the decision is labeled as a good one. If there are topics with lower levels, it is a "bad" decision. If there are no alternatives (all topics have the same annotation level), then it is not a decision at all – no labels are assigned.
- The *metric-B* is a lenient metric. It assumes that as along as the decision made by a student is not the worst, she should be credited for it. As long as there exist a topic with a higher-level annotation (more arrows), the chosen topic constitutes a good navigational decision. Only if the topic is annotated with the maximum-level annotation, and there are topics with lower numbers of arrows, the decision is labeled as a "bad" one.

---

[5] This is a simplification, as students also take into account the goal-based annotation when choosing topics. However, in order to assess the roles of topic-based and goal-based adaptation separately, we have to simplify this analysis.

Both metrics have been computed for all students from the non-adaptive groups (semesters *1&2*) and adaptive group (semesters *3&4*). Several students had to be filtered out from both groups; because, due to the lack of attempts, they have never reached a 1-arrow level of annotation for any of the topics. The final number of students for this analysis is 70: 39 – in the non-adaptive group and 31 – in the adaptive one. Table 4 presents the results of the two *t*-tests. According to both metrics, the adaptive group has made on average significantly more good navigational decisions then the non-adaptive group. The effect size in both cases is very high (*d* ~ 0.8), which underlines the strength of the observed impact of QuizGuide annotations on the students' navigational decisions: students do tend to choose topics with lower-level annotations, thus following the adaptive guidance as intended. This data provides a positive answer to the question we examined in this subsection: *the adaptation does impact students' interaction with the system in a significant way.* Moreover, student behavior was changed in the intended positive direction causing students to focus on topics that they have not sufficiently mastered yet. Thus, despite the observed imperfections on domain and student modeling layers, taken together, the examined layers of adaptation engineering (domain model, student model, adaptation strategy, visual cues) worked as intended affecting student behavior in the desired positive way.

Table 8. Quality of progress-based navigation: comparing the percentage of "good" navigational decisions

	Non-adaptive semesters	Adaptive semesters	Test statistics	Graphical representation
Metric-A (aggressive)	$M = .23$ $Mdn = .18$ $SD = .16$	$M = .39$ $Mdn = .30$ $SD = .26$	$t(68) = 3.14$ $p = .002$ $d = .76$	
Metric-B (lenient)	$M = .55$ $Mdn = .55$ $SD = .17$	$M = .69$ $Mdn = .68$ $SD = .18$	$t(68) = 3.28$ $p = .002$ $d = .79$	

7. Navigational Pattern Analysis

While comparing student work with adaptive and non-adaptive versions, we noticed that typical QuizGuide sessions are both longer and more diverse than QuizPACK sessions (see Table 5). Students attempted more questions through QuizGuide and more frequently accessed material corresponding to different lectures within the same session.

To take a closer look at the nature of these results, we performed a different kind of student activity analysis, taking into account lecture coverage included in all students' actions. Every selection of a question was attributed to the lecture (learning goal) it belongs to. For example, Figure 15 visualizes over 5,500 attempts performed by students using QuizGuide (or QuizPACK) over the entire duration of the "adaptive" Semester 4. Fifteen lectures form the vertical axis. The time of the action is marked on the horizontal axis.

We can detect three zones of activity on this plot. The zone "A" contains the "current" activity that students perform along the lecture stream of the course. It is fairly broad, since homework assignments and in-class quizzes introduce 1-2 weeks delay in shifting the students' focus from previous topics. Zone "B" contains a period of preparation for the final exam. As could be expected, pattern of work with the system is completely different during the exam time: students extensively review the entire course material trying to maximize their overall

knowledge in the domain. Finally, zone "C" contains all actions that students performed during the regular part of the semester, for topics laying far from the "current" lectures. We were particularly interested in this zone. All actions here are not directly motivated by the "current" course situation, but rather initiated by the students themselves, possibly, in an attempt to bridge the existing gaps in the knowledge that should have been acquired earlier.
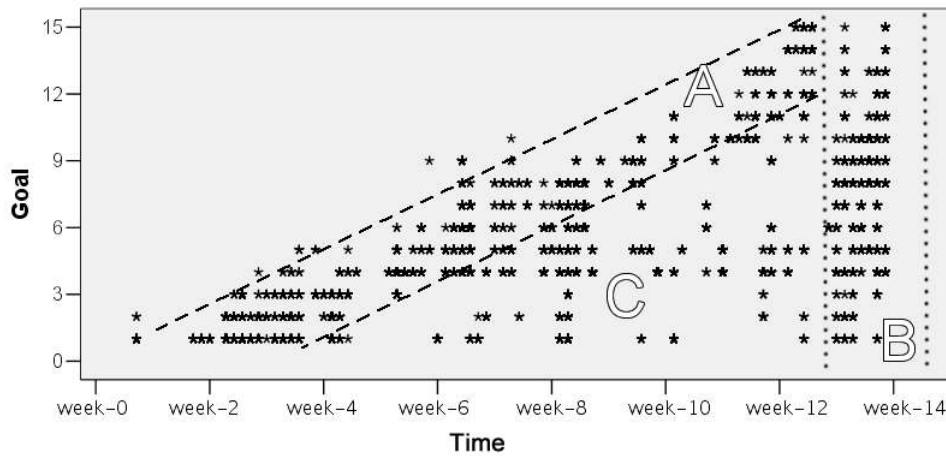

Figure 15. Time distribution of all question answers submitted by students in Semester 4: Zone "A" – lecture stream, zone "B" – final exam cut, and zone "C" – self-regulated extra work with the material of earlier lectures

We used two metrics to assess the intensity of students' self-regulated extra activity:
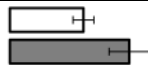- *C-ratio* is the number of attempts in zone "C" made by the student divided by the total number of her attempts;
- *Goal distance* is the average absolute distance between the learning goal, which is current at the time of a question attempt, and the learning goal to which the attempted question belongs (for this metric, we used zones "A" and "C" only).

To evaluate the influence of QuizGuide on self-regulated extra work, we compared the non-adaptive (Semester 1&2) and adaptive (Semesters 3&4) based on these metrics. As shown by Table 9, both measures are significantly higher for the adaptive group. It means, during the semesters when QuizGuide was available, students were more willing to access quizzes lying beyond the current learning goal, in particular, quizzes from uncompleted past topics.

Table 9. Self-regulated extra activity: comparing percentage of students' attempts performed outside the current course focus and the average goal-based distance of students' attempts

	Non-adaptive semesters	Adaptive semesters	Test statistics	Graphical representation
"C" ratio	$M = .20$ $Mdn = .15$ $SD = .26$	$M = .28$ $Mdn = .22$ $SD = .23$	$U = 812$ $p = .026$ $r = .23$	

Goal distance	$M = 5.89$ $Mdn = 4.07$ $SD = 6.48$	$M = 9.56$ $Mdn = 6.04$ $SD = 9.82$	$U = 811$ $p = .027$ $r = .23$	

The observed effect can be explained by the fact that progress-based and prerequisite-based adaptive annotations generated by QuizGuide directed students' attention to the material related to those earlier lectures, which were not understood and not sufficiently explored. The main stream of activity stays in the zone of recommended work (zone "A" on Figure 15), which is moving along *class progress* – topics are annotated as ready and current in the week they are presented. However, QuizGuide provides clear guidance to the past topics that are important to understand the current topics, but not have not yet been mastered: target color indicates the current topics prerequisites and the lack of arrows shows the insufficient level of knowledge. With this guidance, students struggling with a current topic can easily focus on its prerequisites that have not been well understood. The progress-based annotation can further motivate students to review the previous topics and earn annotations of higher levels. Thus, the navigation pattern analysis provides some evidence that a combination of progress-based and prerequisite-based adaptive annotations encourages students to return to past topics and to bridge the knowledge gap required to work successfully with the current ones.

8. Evaluation of QuizGuide by Students

Evaluation of an adaptive system by its users is a popular approach for assessing the perceived value of the system. Such evaluations are usually organized after the users completed the designated task, and take the form of questionnaires testing users' opinion about different features of the system. The questionnaires need to be carefully designed in order for every question to elicit the intended piece of users' opinion, and only one at a time. The results of such evaluation sometimes help validate the outcomes of other tests. In the case of QuizGuide, the positive attitude of students was one of the factors facilitating the success of the system.

At the end of each semester, students were invited to take a set of questionnaires that asked them about various aspects of the course experience, including self-assessment quizzes. In Semester 3, once QuizGuide was introduced, we extended the questionnaire with new questions. More QuizGuide specific questions were added in Semester 4, after we processed the first dataset. Questions were designed using a five-point Likert scale.

Answering the questionnaires was voluntary; therefore, only a part of students provided us with their opinions about the system. Furthermore, to ensure the reliability of their judgments, for the analysis we selected only answers from the students who used self-assessment quizzes sufficiently (more than 30 questions answered over the semester) and who tried the adaptive version at least once. The final dataset included answers from 19 students in Semester 3 and 13 students in Semester 4.

Table 10 shows the set of QuizGuide-related questions together with the results of one-sample Wilcoxon signed rank test[6] verifying whether an average student answer significantly differs from the neutral opinion. Questions 1-3 were available to the students of both Semesters 3 and 4; these questions estimated student attitude toward general aspects of the system. Questions

---

[6] Student responses were not normally distributed, in addition, it is recommended in the literature to avoid using parametric statistical testing with ordinal data; therefore, we could not apply one-sample t-test.

4-8 were added only during the last semester; this set of questions asked student what they think about the adaptation mechanism of QuizGuide. As can be seen from the table, for all eight questions, student opinion about different features of QuizGuide was significantly positive (an average score was significantly greater then the neutral "*No strong opinion*" option).

Table 10. Student answer to the questionnaires evaluating their opinion about QuizGuide.

N	Question	Answers (0-4)	Test statistics	Graphical representation
1	Self-assessment quizzes served by QuizPACK/QuizGuide helped me during the course	$M = 3.34$ $Mdn = 3.00$ $SD = .77$	$U = 371$ $p < .001$ $r = .88$	
2	I preferred accessing self-assessment quizzes with adaptive guidance (through QuizGuide)	$M = 3.07$ $Mdn = 3.00$ $SD = .80$	$U = 292$ $p < .001$ $r = .83$	
3	I like the interface of QuizGuide	$M = 3.21$ $Mdn = 3.00$ $SD = .62$	$U = 351$ $p < .001$ $r = .91$	
4	I think the target-arrows icons in QuizGuide are informative	$M = 3.45$ $Mdn = 3.0$ $SD = .52$	$U = 66$ $p = .001$ $r = .91$	
5	I benefitted from adaptive annotation of topics in QuizGuide	$M = 3.09$ $Mdn = 3.0$ $SD = .83$	$U = 62$ $p = .005$ $r = .80$	
6	Usually, I tried to pay more attention to the topics where I had fewer arrows and took quizzes belonging to those topics	$M = 2.73$ $Mdn = 3.0$ $SD = .79$	$U = 41$ $p = .012$ $r = .70$	
7	Usually, I followed QuizGuide suggestions of the most relevant topic and took quizzes for the topic with an icon of the most intense color	$M = 2.82$ $Mdn = 3.0$ $SD = .98$	$U = 33$ $p = .018$ $r = .65$	
8	Usually, the number of arrows on the target icon correctly reflected the level of my knowledge for the corresponding topic	$M = 2.82$ $Mdn = 3.0$ $SD = .87$	$U = 41$ $p = .012$ $r = .70$	

Figure 16 visualizes the ratios of student answers to the same set of QuizGuide-related questionnaire items. The percentage of positive or strongly positive responses ranges from 64% (question 7) to 100% (question 4). The percentage of negative responses ranges from 0% (questions 3 and 4) to 9% (questions 5-8). No single strongly negative opinion was reported for any of the questions.

Peter Brusilovsky 7/27/2016 10:35 AM
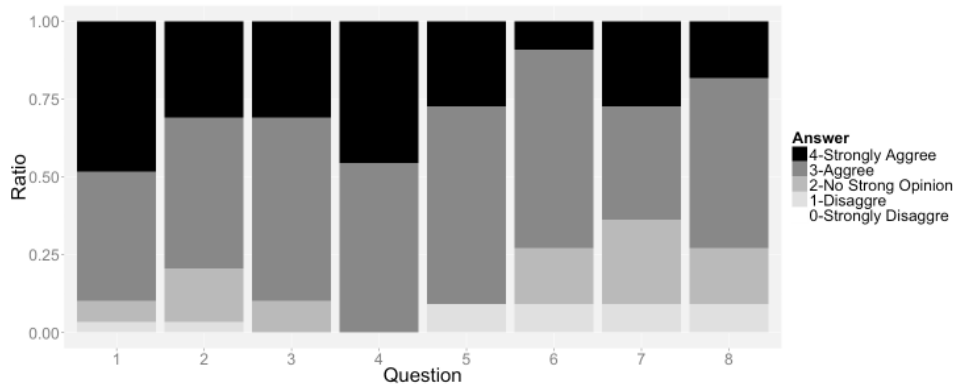**Deleted:** Figure 16

Figure 16. Students' subjective evaluation of QuizGuide topic-based adaptation.

As the data show, 90% of students agree that work with self-assessment quizzes was helpful for them (Question 1). This nicely correlates with our previous findings that demonstrated several significant effects of self-assessment quizzes served by QuizPACK on a range of course performance parameters (Brusilovsky,Sosnovsky 2005b).

Questions 2 and 3 assessed student perception of QuizGuide as a whole. Question 2 reveals that 79% of students preferred accessing self-assessment quizzes through adaptive QuizGuide. This number is a little higher than the objective usage parameters show. Question 3 shows that the interface of QuizGuide was appreciated by 90% of students despite (or, possibly, thanks to) its simplicity.

The rest of the questions show that student attitude to the most critical features of QuizGuide was also very positive. All students (100%) agree that the target-arrows icons in QuizGuide are informative (Question 4) and 91% of students agree that adaptive annotation of topics with icons is helpful (Question 5). At the same time, while recognizing the value of adaptive annotations, the students do not always follow it as the system suggests. Only 73% agree that adaptive annotation causes them to take more quizzes for the topics with fewer arrows, i.e. follow progress-based navigation (Question 6); and only 64% agree that it causes them to take more quizzes for the current topics with bright blue icons, i.e. follow immediate goal-based navigation (Question 7). It indicates that some fraction of students prefer to make their own navigation choices while appreciating QuizGuide navigation support as informative and valuable. This data stresses the importance of flexibility provided by adaptive navigation support – it provides personalized guidance while leaving it to the students to make navigation decisions.

The last questionnaire item (Question 8) is especially interesting to us, as an indicator of student opinion on the quality of topic-based UM. 73% of subjects believe that, usually, the number of arrows on the target icon correctly reflects the level of their knowledge for the corresponding topic. This means that, in spite of the fact that the objective analysis shows relatively low predictive validity of the implemented UM mechanism, students do not feel that the system misinterprets their knowledge. An even more surprising data comes from the students' answers to the open question in the questionnaire. When asked to indicate the most problematic feature of QuizGuide, students have been provided with a set of options and also given the opportunity to suggest their own feature. Among the predefined options, there have

been the following two: "*The system is too strict; I felt that I knew the material good enough before QuizGuide gave me 3 arrows*" and "*The system is too lenient. I felt that I got 3 arrows before I really understood the topic*". None of the students marked the "strict" option as a concern while four out of eleven students have indicated *leniency* as one of the problematic features of QuizGuide. It shows that students' concerns about student modeling precision were rather opposite to our own concerns. Recall that in section 6.2 we discussed that the in contrast to BKT, our UM formula often underestimates the level of knowledge students have achieved (Figure 13), and, as a result, mastered topics tend to be annotated with lower-level knowledge icons. Yet, several students felt the opposite – they believed that the system was giving them higher-level annotations before they mastered the corresponding topics. Thus the tendency of our "teacher-inspired" modeling approach to underestimate student knowledge in comparison with BKT (that tends to overestimate it) might be a positive rather than a negative feature in the context of adaptive navigation support. It further stressed the importance to assess not just the quality of student models, but student perception of adaptation produced using this model.

Overall, as the data shows, the students regarded the system and its main personalization features very positively, which is important from both conceptual and practical perspectives.

9. Discussion

The QuizGuide system is an attempt to build a practical educational solution that supports user-adapted access to educational content in the context of a real semester-long course. On the teacher side, QuizGuide relies on a minimalist approach to domain, student and content modeling, thus enabling easy and familiar authoring procedure. On the student side, its interface combines adaptive navigation support and open student modeling to attract student attention to the topics important at the moment. A straightforward personalization design applied in QuizGuide made it possible to implement adaptation layer as a lightweight value-added service, largely independent of the content being adapted. The practical value and a relative simplicity of implementation made this approach very attractive for extensive follow-up work. After its pilot in QuizGuide system for C programming, the approach has been successfully employed in other domains such as SQL (Brusilovsky et al. 2010) and Java (Hsiao et al. 2010) and embedded as one of the key approaches in generic navigation support systems (Yudelson 2010). This paper, however, presents the first attempt to look much deeper into this approach and applies a large arsenal of evaluation techniques to analyze it from multiple perspectives including:
- the added value of adaptation (Section 5),
- the quality of the models contributing to adaptation (Section 6),
- the navigational behavior exhibited by the students in the presence of adaptation (Section 7) and
- the aggregated subjective opinion that student form about the system (Section 8).

9.1. QUIZGUIDE TOPICS: KNOWLEDGE COMPONENTS WITH SPECIAL FEATURES

Similar to finer-grained concepts, the main purpose of coarse-grained topics used by QuizGuide is to represent elements of knowledge in the domain of discourse to serve as a basis for identification of the student's understanding of corresponding learning material, and to support proper adaptation. However, the topic-based approach to knowledge modeling is different in several ways:

– Topics provide a means for learning material *aggregation*, instead of traditional *indexing*. As a result, the relationships between topics and pieces of learning content are "1-to-many" (e.g. one topic corresponds to many quizzes/questions) and topic-based "indexes" do not exceed manually manageable numbers.

– Topics provide a natural approach for a classroom teacher to organize the course into logically separate units and assign appropriate pieces of content to them. Consequently, a classroom teacher can easily author a topic-based domain model while developing a course structure.

– In adaptive systems, topics can play two roles: as knowledge components for the student modeling and content-based adaptation, and as interface elements for the content structuring and navigation. As the interface elements, topics allow the system to communicate to a student the organization of the domain, as well as the current state of the student model in familiar terms (the set, order, and titles of the topics reflect the structure and vocabulary of the course defined by the teacher), thus enabling meaningful knowledge-based navigation through the content.

– Topics are coarse-grained; therefore, when relying on topic-based knowledge modeling, some sacrifice in the model precision is being made. At the same time, QuizGuide demonstrates that even such a coarse-grain approach to domain and student modeling can result in efficient adaptive interfaces that have a positive impact on several aspects of learning.

– Topics are subjective. If comparing topic structures of the same course developed by different teachers, one can expect them to be quite different. The presence or absence of a single topic, the naming labels, the size of particular topics, the inter-topics relations, and the scope of the entire set can vary from one structure to another based on the personal decision made by a teacher.

In summary, topics are unique knowledge components, which in the framework of adaptive learning, have some positive and negative features compared to smaller concepts. An important advantage of a topic is that, while ensuring sufficient adaptation quality, it provides the classroom teacher with a natural approach to adaptive content authoring and student progress tracking.

9.2. EVALUATION RESULTS: THE ADDED VALUE OF ADAPTATION

The most straightforward way to evaluate the added value of adaptation is to compare the adaptive version of the system with the non-adaptive one. Most often, such setup is implemented by turning off the adaptive functionality of the developed adaptive system. De Bra (2000) criticizes this approach, for being "not a fair comparison". Indeed, the non-adaptive counterpart is often at disadvantage, since the original design of the system assumes adaptive functionality. In our case, such consideration does not hold, as the QuizPACK system providing the content was originally designed as a full-functioning application. The adaptive version of the evaluation setup consists of the QuizPACK-QuizGuide tandem, where QuizPACK serves the same set of questions/quizzes as in the non-adaptive mode and QuizGuide provides the adaptive navigation support.

We have observed several effects of exposing students working with self-assessment quizzes to the topic-based adaptive navigation of QuizGuide:

- *Knowledge gain*. On both metrics (normalized/absolute), work with QuizGuide resulted in significantly higher knowledge gain. An average student using adaptive interface for

accessing quizzes learnt about .5 sigma more than an average student using a non-adaptive system. An average "adaptive" student is above 70% of the "non-adaptive" group, when it comes to knowledge gain.

- *Motivation and engagement*. Students have worked with the self-assessment quizzes significantly more during "adaptive" semesters compared to the "non-adaptive" ones. This increase in motivation has been observed in terms of a multitude of parameters: they answered twice as many questions, covered twice as many quizzes and registered 50% more sessions. The sessions themselves became 50% lengthier (in terms of attempts) and almost twice more diverse (in terms of quizzes and questions). The magnitude of the increase is considerably large; therefore, the motivation effect is likely to be one of the reasons for the learning effect.
- *Within system performance*. Even though we did not register any difference between the adaptive and non-adaptive groups in terms of success ratio, we have found a significant effect of adaptation when compared students who primarily used QuizGuide and those who mostly accessed quizzes in the non-adaptive manner. QuizGuide did help students to choose topics more wisely and achieve a better success rate of their attempts.
- *Global navigational patterns*. Another type of analysis that we performed looked closer at the course-wide picture of the students' access behavior with and without adaptation. It turns out that in the presence of adaptation, students tend to work significantly more with the quizzes that lay behind the current topic (as per the course schedule). This navigational activity represents self-regulated extra work with the past learning material. QuizGuide uses several mechanisms to engage students into such activity. The lack of arrows on a topic icon and the special coloring of pre-requisite topic icons can inspire a student to spend more time with the relevant quizzes, even if the topic no longer belongs to the current learning goal of the course.

9.3. EVALUATION RESULTS: LAYERED EVALUATION

The topic-based adaptation of QuizGuide has had a number of positive effects on student learning and performance both within and outside the system. However, the unconventional simplified approach to domain, content and student modeling implemented in QuizGuide prompted for a deeper look into the quality of the models underlying adaptation, and the adaption method itself. The student logs collected in the system over the four consecutive semesters provided us with enough evidence to perform a set of educational data mining analyses of the three layers of QuizGuide personalization approach:

- *Topic-based domain modeling*. From the point of adaptive educational research, the quality of a domain model is defined by its ability to support meaningful modeling of the learning process. When student mastery measured in terms of the elements of a domain model grows according to the power law of practice, one can say such a model represents the units of domain semantics on the psychologically appropriate level. The aggregate learning curves for QuizGuide topics demonstrate relatively poor fit by power function: $R^2 = .39$. The follow-up analysis has shown that by reducing the size of the assessment unit we can achieve a much better fit ($R^2 = .70$); similarly, by filtering out too difficult and too simple questions we also brought the topic-based learning curve much closer to the power function ($R^2 = .67$).
- *Topic-based student modeling*. We evaluated the knowledge modeling formula used by QuizGuide from the point of its predictive validity. In other words, we examined

how successfully the formula can predict the result of a student answer to the next question give a stream of his previous attempts to the questions of the same topic. We contrasted the results produced by the QuizGuide formal against Bayesian knowledge tracing (BKT), which is the most popular approach for student modeling. The QuizGuide formula has shown a significant effect when compared to the random prediction; however, it was also significantly outperformed by BKT. The most important problem of QuizGuide knowledge modeling method was its conservative nature toward acknowledging student progress. It was designed to simulate a teacher view on student performance, and required students to prove themselves on a substantial amount of questions constituting a topic before treating the topic as mastered.

- *Topic-based adaptive guidance*. The main goal of the adaptation approach implemented in QuizGuide is to inform students about the current importance of topics and the utility of corresponding quizzes for their learning. The final decision about which topic to focus on lies with the students, as QuizGuide does not restrict their access in any way. Therefore, when evaluating the quality of QuizGuide adaptation, we tried to estimate how well the system supports them in this decision making and, consequently, how often they follow system's suggestions. The analysis shows that, in the presence of adaptation student are making significantly more (about 15%) "good" navigational decisions, i.e., they focus significantly more on the topics that require additional work.

9.4. Evaluation Results: Evaluation by Students

In order to elicit the perceived value of QuizGuide, at the end of each semester, we administered a questionnaire asking students for their opinion on different features of QuizGuide. The overall result of this analysis is that, on every question, an average student answer was significantly positive. Students felt that QuizGuide helped them during the semester. They followed both the progress-based and the goal-based navigation support; and they believed that the adaptive annotations were informative and beneficial and, what especially important, correctly reflected their knowledge for the corresponding topics. A somewhat surprising result came from student answers to the open question: none of the students reported that QuizGuide underestimated their knowledge; on the contrary, several students though that, sometimes, QuizGuide awarded them with the high-level progress-based annotations too soon. This feedback provides considerable evidence that, student modeling and adaption in QuizGuide was perceived as appropriate and that a conservative nature of the "teacher-inspired" conservative modeling approach might be not a weakness, but rather a strength, in the explored context.

9.5. Summary

The layered analysis of the QuizGuide topic-based personalization mechanism has shown that the selected topics do not support a high precision of the domain modeling (Section 6.1) and that the student modeling mechanism used in QuizGuide tends to underestimate student ability to solve questions (Section 6.2). Nevertheless, students were able to use QuizGuide effectively. They follow adaptive cues (Sections 6.3 and 7); they access more diverse set of topics (Sections 7 and 5.3). In the presence of QuizGuide adaptation, they work more actively (Section 5.3), diversely (Section 5.3) and meaningfully (Section 5.2), and they learn significantly more

(Section 5.1).They trust topic-based modeling and guidance and believe that they benefit from it (8). Thus, from the pragmatic perspective, the adaptive system QuizGuide based on coarse-grained topics has performed extremely well.

This leaves us with two interesting questions to conclude the discussion.

- Can the topic potential as modeling units be improved?
- How does QuizGuide provide high-quality adaptation based on low-quality knowledge models?

Improving modeling quality. There are two main sources of problems with QuizGuide topics as domain modeling units: topics covering too much learning material and questions with extreme difficulty values. More careful selection of topics that are more cohesive and smaller in size will improve the quality of a topic based domain model and the quality of knowledge assessment it can support. The problem of too difficult and too easy questions can be addressed either by designing questions with better assessment parameters, or by more careful manipulation of existing questions in order to match the current student ability and the difficulty of the questions. For example, questions within a topic could be ordered from easy to difficult during the authoring phase. Another solution could be an adaptive mechanism for selecting questions with the appropriate difficulty or navigating students to the right question within a topic (see (Sosnovsky et al. 2008) for an example). Finally, the system can simply inform students about question characteristics and rely on their judgment for proper question selection.

From modeling to adaptation. When a student modeling component of an AES has problems with predicting the current state of learning, the effectiveness of the entire system is at risk. However, an important question to consider is – which prediction errors are more problematic in a given context. While BKT has proven to be more accurate than the QuizGuide formula, both methods have problems with predicting student ability to solve the next question, and the majority of their incorrect predictions lie on the opposite sides of the "perfect model". QuizGuide formula has a higher miss rate (false negative rate) due to consistent *underestimation* of student ability. BKT, in its turn, has a higher fall-out (false positive rate) by overestimating student topic mastery. When these two types of errors are translated into adaptive navigation support decision, QuizGuide formula encourages students working more on the topics that they could have abandoned already, while BKT formula guides students away from the topics that they have not mastered yet. We believe that, in the adaptive hypermedia setting settings, the first type of modeling error is less dangerous. Moreover, the adaptive navigation support applied in QuizGuide was designed to match the conservative teacher-inspired approach. When translating the knowledge levels into navigational cues, QuizGuide compensated for the strict nature of its modeling formula by annotating any topic with a knowledge level higher than .5 with a 3-arrows icon.

As a result, QuizGuide performed its task to guide students within a body of non-mandatory educational content really well, with no negative effects of its conservative students modeling observed. Indeed, a potential drawback of unnecessary conservative progress annotation is that the students will have to go through redundant drilling, become bored and withdraw from the system. The in-class evaluation of QuizGuide has clearly shown that the student motivation was, in fact, remarkably increased. Moreover, the "playing it safe" approach of QuizGuide modeling formula might have been one of the reasons for students to stay with the system longer and achieve better results. In addition, during the subjective evaluation of the system, none of the students reported that the system underestimates their knowledge; in fact, some of them felt the opposite. Hence, while acknowledging that the original QuizGuide

modeling approach requires an improvement, we think that in the explored context, it served rather well. In general, an important argument can be made based on these results: while the quality of UM is among the critical factors for any adaptive application, on the practical level, it matters only when considered together with the adaptation approach employing the results of UM and translating them into adaptive decisions communicated to the users of the application.

Specifically, the topic-based adaptive navigation support technology used in QuizGuide appeared to be a good match for the topic-based modeling approach. Adaptive navigation support is among the most flexible adaptation technologies. Unlike intelligent turning systems that seek to control interaction between the student and the learning material very closely and leave little freedom to the student, educational systems employing adaptive navigation support, do not restrict students access, but rather inform them about the next best move. Thus, accuracy of student modeling is less critical for them. It may be even an effective strategy to react to student progress with some delay, as QuizGuide does. In contrast, with the mastery-learning adaptation approach, for which BKT was originally developed, higher precision of BKT could be more important and its tendency to move students ahead faster than they deserve might prevent student frustration in this context where students can't advance without system permission.

10. Overview of Student Modeling in Adaptive Educational Systems

AESs belong to the class of *user-adaptive software systems* (Schneider-Hufschmidt et al. 1993). The key component of all systems in the class is the user model – a representation of information about an individual user that is essential for an adaptive system to provide *the adaptation effect*, i.e., to behave differently for different users. As a class, adaptive systems explored how to model a broad range of user features to deliver adaptation (Brusilovsky,Millán 2007); however, each specific category of adaptive systems focus on the features that are most essential for this category to adapt to. For example, adaptive help systems focused on goal modeling, adaptive news systems explored interest modeling, and recommender systems perfected user taste modeling. In this spectrum, AESs achieved most of their known success by modeling user knowledge (Brusilovsky,Millán 2007). Moreover, this category of systems was both the first (Carbonell 1970) and the most successful in modeling user knowledge. As a result, models of user knowledge are frequently called student models, following the name that was originally used for this kind of models in the field of AESs (Self 1974).

Surprisingly, the variety of student model types in the field is rather small. Early adaptive systems explored a range of interesting ideas from bug libraries (Brown,Burton 1978) to genetic graphs (Goldstein 1979), however, gradually the *overlay student modeling* approach originally suggested in (Goldstein,Carr 1977) emerged as most popular due its optimal combination of power and simplicity. Nowadays, the dominated majority of AESs use some form of overlay modeling.

The key idea of the overlay student modeling is representing student knowledge as an overlay of a structured domain model, which, in turn, is perceived as a model of an "ideal student" or expert. A structured domain model decomposes the body of knowledge about the domain into a set of knowledge items (KI). These items can be named differently in different sources (rules, concepts, knowledge units, topics, knowledge elements, learning objectives, learning outcomes), but in all the cases they denote elementary fragments of domain knowledge or information. Depending on the domain, the application area, the nature of modeled knowledge, and the choice of the designer, KIs can represent larger or smaller fragments of

knowledge: from a relatively large chunk of knowledge (Papanikolaou et al. 2003) down to elementary facts (Oberlander et al. 1998; Carbonell 1970), rules (Corbett,Anderson 1995; Ritter 1997), or constraints (Ohlsson 1992; Mitrovic 2003). In some systems, KIs could be connected by pedagogical relationships such as prerequisite (Henze et al. 1999; Weber et al. 2002; Papanikolaou et al. 2003) or semantic relationships such as "is-a" and "part-of" (Brusilovsky,Cooper 2002; De Bra,Ruiter 2001; Vassileva 1998). These relationships help improving both student modeling and personalization. In this context, QuizGuide uses a coarse-grain domain model with 22 KIs called topics and prerequisite relationships between them (Figure 7).

Representing student knowledge as an overlay of the domain model means that each individual student model maintains a separate estimation student knowledge/mastery level for each domain model KI. While the classic overlay models used a binary value (known – not known) to represent knowledge of each KI, modern systems used weighted overlay models that can distinguish several levels of user's knowledge about each KI. Historically, there are three popular forms of weighted overlay models: qualitative, simple numeric, and uncertainty-based that correspond to three approaches to UM and adaptation. Qualitative models represent knowledge of a KI as a qualitative value such as good-average-poor and are used mostly by systems with rule-based student modeling and adaptation components (Papanikolaou et al. 2003; Brusilovsky,Anderson 1998). Simple numeric models use quantitative estimates (for example, from 0 to 100) to represent the level of student knowledge (Brusilovsky et al. 1998; De Bra,Ruiter 2001). These models are explored by systems with simple algebraic approaches to knowledge modeling and propagation. The uncertainty-based models use different forms of uncertainty management such as Bayesian networks of fuzzy logics to model student knowledge. The user knowledge in these models are represented in the form dictated by the selected approach – most frequently, a probability that the user knows the KI (Henze,Nejdl 1999; Corbett,Anderson 1995; Conati et al. 2002). The QuizGuide student model presented in this paper is a good example of a simple numeric model, while the BKT approach used for comparison in Section 6.2 is a classic example of probabilistic student modeling (Corbett,Anderson 1995).

Over the last 15 years, uncertainty-based models gradually evolved from being a curiosity to a dominated student modeling approach. This shift was caused by the increased availability of student log data that allowed data-driven tuning of various modeling parameters, which was the critical step to increase the quality of the process. Just like the overlay model evolved as a golden standard due to its combination of power and simplicity, BKT approach evolved as the most popular approach for probabilistic student modeling. While being less sophisticated that some advanced approaches such as Bayesian networks (Conati et al. 2002), it combines expressive power with an ability to learn from data. Being originally developed for modeling of student procedural knowledge of LISP, it has been successfully used for skill modeling in many domains from algebra (Koedinger et al. 1997) to genetics (Baker et al. 2011) and was more recently applied for modeling conceptual knowledge (Pirolli,Kairam 2013). Over the last 5 years, a number of interesting extensions over the original BKT were suggested to address some of the known problems of this approach (Baker et al. 2010; Pardos,Heffernan 2011; González-Brenes et al. 2014; Falakmasir et al. 2013), which, in turn, made this family of approaches more popular.

Peter Brusilovsky 7/27/2016 10:35 AM
Deleted: Figure 7

11. Conclusion and Future Work

This paper presents the first attempt at in-depth investigation of topic-based personalization in the context of an educational system. We introduced the topic-based approach for student modeling and adaptation and presented its implementation in QuizGuide system. We also performed an extensive analysis of this implementation within several semester-long courses. The multiple evaluation techniques that we applied provided valuable data about the strengths and weaknesses of this approach and led to a number of interesting discussion points. We believe that the immediate results of our work, the variety of evaluation approaches that we applied, and the provided discussion could be valuable for the designers of practical AESs. At the same time, despite considerable volume of design and evaluation work, this research has a number of limitations that we will try to address in our future exploration of the topic-based approach.

Broadening the scope. Although, the amount of the collected data used in this study is substantial, we recognize that it is designed around one subject domain, one set of topics, one type of content, one system and one adaptation technique. In order to draw more general and far-reaching conclusions about the potentials of topic-based personalization, additional work needs to be conducted. We have started these efforts by implementing similar topic-based adaptation in two other programming domains (Java and SQL), and achieving similar learning and motivation effects.

Assessing the authoring side. One of the main premises of the topic-based approach is its simplicity. Design of knowledge models based on aggregation of learning material into coarse-grained and subjective topics is a relatively easy and familiar procedure for an average classroom teacher. However, in order to verify that an unprepared teacher can create a functional topic-based model, a teacher-oriented authoring system for assembling topic-based adaptive courses should be built and a set of studies with third-party teachers should be conducted.

In vitro versus in vivo evaluation. Running real classroom studies was important for us in order to assess the practical, ecologically valid impact of the topic-based personalization. Yet, the nature of the classroom study makes it hard to assess multiple individual factors that contribute to our approach. We believe that a set of controlled lab studies should be used to perform more accurate evaluation of the impact of the suggested technology as well as to compare different design options (such as different interface organization, different visual cues, etc.).

With all these limitations, we believe that the amount of collected evidence is sufficient to recommend the topic-based adaptation approach for a broader use and hope that our paper provides sufficient guidance to those who would like to explore it in other contexts and fields.

12. Acknowledgments

References

Uncategorized References

Ahn, J.-w., Brusilovsky, P., He, D., Grady, J., Li, Q.: Personalized Web Exploration with Task Models. In: the 17th international conference on World Wide Web, WWW '08, Beijing, China, April 21-25, 2008 2008, pp. 1-10. ACM

Anderson, J.R., Corbett, A.T., Koedinger, K.R., Pelletier, R.: Cognitive tutors: Lessons learned. The Journal of the Learning Sciences **4**(2), 167-207 (1995)

Apted, T., Kay, J., Lum, A.: Supporting metadata creation with an ontology built from an extensible dictionary. In: De Bra, P., Nejdl, W. (eds.) Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004), Berlin, August 23-26, 2004 2004. Lecture Notes in Computer Science, pp. 4-13. Springer-Verlag

Aroyo, L., Dicheva, D.A.: Concept-based approach to support learning in a Web-based support environment. In: Moore, J.D., Redfield, C.L., Johnson, W.L. (eds.) AI-ED'2001, Amsterdam 2001, pp. 1-12. IOS Press

Baker, R.S., Corbett, A.T., Aleven, V.: More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In: Woolf, B., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) the 9th international conference on Intelligent Tutoring Systems, Montreal, Canada, , June 23-27, 2008 2008, pp. 406-415. Springer Berlin Heidelberg

Baker, R.S.J.d., Corbett, A.T., Gowda, S.M., Wagner, A.Z., MacLaren, B.A., Kauffman, L.R., Mitchell, A.P., Giguere, S.: Contextual Slip and Prediction of Student Performance After Use of an Intelligent Tutor. In: De Bra, P., Kobsa, A., Chin, D. (eds.) 18th International Conference on User Modeling, Adaptation, and Personalization (UMAP 2009), Big Island, HI, USA, June 22-24, 2010 2010. Lecture Notes in Computer Science, pp. 52-63. Springer

Baker, R.S.J.d., Pardos, Z.A., Gowda, S.M., Nooraei, B.B., Heffernan, N.T.: Ensembling Predictions of Student Knowledge within Intelligent Tutoring Systems. In: 19th International Conference on User Modeling, Adaptation, and Personalization, UMAP 2011, Girona, Spain, July 11-15, 2011 2011, pp. 13-24. Springer-Verlag

Beck, J., Chang, K.-m., Mostow, J., Corbett, A.: Does Help Help? Introducing the Bayesian Evaluation and Assessment Methodology. In: Woolf, B., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) the 9th international conference on Intelligent Tutoring Systems, Montreal, Canada, , June 23-27, 2008 2008, pp. 383-394. Springer Berlin Heidelberg

Brown, J.S., Burton, R.R.: Diagnostic models for procedural bugs in basic mathematical skills. Cogn. Science **2**, 155-192 (1978)

Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Modeling and User-Adapted Interaction **6**(2-3), 87-129 (1996)

Brusilovsky, P.: Adaptive hypermedia. User Mod. & User-Adapt. Interact. **11**(1/2), 87-110 (2001)

Brusilovsky, P.: Developing Adaptive Educational Hypermedia Systems: From Design Models to Authoring Tools. In: Murray, T., Blessing, S., Ainsworth, S. (eds.) Authoring Tools for Advanced Technology Learning Environments: Toward cost-effective adaptive, interactive, and intelligent educational software. pp. 377-409. Dordrecht, Kluwer (2003)

Brusilovsky, P.: KnowledgeTree: A distributed architecture for adaptive e-learning. In: 13th International World Wide Web Conference, WWW 2004, New York, NY, 17-22 May, 2004 2004, pp. 104-113. ACM Press

Brusilovsky, P.: Adaptive navigation support. In: Brusilovsky, P., Kobsa, A., Neidl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization, vol. 4321. Lecture

Notes in Computer Science, pp. 263-290. Springer-Verlag, Berlin Heidelberg New York (2007)

Brusilovsky, P.: Adaptive Hypermedia for Education and Training. In: Durlach, P., Lesgold, A. (eds.) Adaptive Technologies for Training and Education. pp. 46-68. Cambridge University Press, Cambridge (2012)

Brusilovsky, P., Anderson, J.: ACT-R electronic bookshelf: An adaptive system for learning cognitive psychology on the Web. In: Maurer, H., Olson, R.G. (eds.) WebNet'98, World Conference of the WWW, Internet, and Intranet, Orlando, FL, November 7-12, 1998 1998, pp. 92-97. AACE

Brusilovsky, P., Cooper, D.W.: Domain, Task, and User Models for an Adaptive Hypermedia Performance Support System. In: Gil, Y., Leake, D.B. (eds.) 2002 International Conference on Intelligent User Interfaces, San Francisco, CA, January 13-16, 2002 2002, pp. 23-30. ACM Press

Brusilovsky, P., Eklund, J.: A study of user-model based link annotation in educational hypermedia. J. Universal Comp. Science **4**(4), 429-448 (1998)

Brusilovsky, P., Eklund, J., Schwarz, E.: Web-based education for all: A tool for developing adaptive courseware. In: Ashman, H., Thistewaite, P. (eds.) Seventh International World Wide Web Conference, Brisbane, Australia, 14-18 April 1998 1998, vol. 1-7, pp. 291-300. Elsevier Science B. V.

Brusilovsky, P., Karagiannidis, C., Sampson, D.: Layered evaluation of adaptive learning systems. Int. J. Cont. Eng. Education & Lifelong Learning **14**(4/5), 402 - 421 (2004)

Brusilovsky, P., Knapp, J., Gamper, J.: Supporting teachers as content authors in intelligent educational systems. Int. J. Knowledge & Learning **2**(3/4), 191-215 (2006)

Brusilovsky, P., Millán, E.: User models for adaptive hypermedia and adaptive educational systems. In: Brusilovsky, P., Kobsa, A., Neidl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization, vol. 4321. Lecture Notes in Computer Science, pp. 3-53. Springer-Verlag, Berlin Heidelberg New York (2007)

Brusilovsky, P., Miller, P.: Course Delivery Systems for the Virtual University. In: Tschang, T., Della Senta, T. (eds.) Access to Knowledge: New Information Technologies and the Emergence of the Virtual University. pp. 167-206. Elsevier Science, Amsterdam (2001)

Brusilovsky, P., Sosnovsky, S.: Engaging students to work with self-assessment questions: A study of two approaches. In: 10th Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE'2005, Monte de Caparica, Portugal, June 27-29, 2005 2005a, pp. 251-255. ACM Press

Brusilovsky, P., Sosnovsky, S.: Individualized Exercises for Self-Assessment of Programming Knowledge: An Evaluation of QuizPACK. ACM J. Educ. Res. in Comp. **5**(3), Article No. 6 (2005b)

Brusilovsky, P., Sosnovsky, S., Lee, D., Yudelson, M., Zadorozhny, V., Zhou, X.: Learning SQL programming with interactive tools: from integration to personalization. ACM Transactions on Computing Education **9**(4), Article No. 19, pp. 11-15 (2010).

Brusilovsky, P., Sosnovsky, S., Shcherbinina, O.: User Modeling in a Distributed E-Learning Architecture. In: Ardissono, L., Brna, P., Mitrovic, A. (eds.) 10th International User Modeling Conference, Berlin, July 24-29, 2005 2005a. Lecture Notes in Artificial Intelligence, pp. 387-391. Springer Verlag

Brusilovsky, P., Sosnovsky, S., Yudelson, M.: Addictive links: The motivational value of adaptive link annotation. New Rev. Hypermedia & Multimedia **15**(1), 97-118 (2009)

Brusilovsky, P., Sosnovsky, S., Yudelson, M., Chavan, G.: Interactive Authoring Support for Adaptive Educational Systems. In: Looi, C.-K., McCalla, G., Bredeweg, B., Breuker, J. (eds.) 12th International Conference on Artificial Intelligence in Education, AIED'2005, Amsterdam, July 18-22, 2005 2005b, pp. 96-103. IOS Press

Bull, S.: Supporting learning with open learner models. In: 4th Hellenic Conference on Information and Communication Technologies in Education, Athens, Greece, September 29 - October 3, 2004 2004, pp. 47-61

Carbonell, J.R.: AI in CAI: An artificial intelligence approach to computer aided instruction. IEEE Trans. Man-Machine Syst. **MMS-11**(4), 190-202 (1970)

Carmona, C., Bueno, D., Guzmán, E., Conejo, R.: SIGUE: Making Web Courses Adaptive. In: De Bra, P., Brusilovsky, P., Conejo, R. (eds.) Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2002), Málaga, Spain, May 29-31, 2002 2002. Lecture Notes in Computer Science, pp. 376-379. Springer-Verlag

Chin, D.: Empirical Evaluations of User Models and User-Adapted Systems. User Modeling and User-Adapted Interaction **11**(1-2), 181-194 (2001)

Conati, C., Gertner, A., Vanlehn, K.: Using Bayesian Networks to Manage Uncertainty in Student Modeling. User Modeling and User-Adapted Interaction **12**(4), 371-417 (2002)

Conejo, R., Guzman, E., Millán, E.: SIETTE: A Web-based tool for adaptive teaching. Int. J. AI in Education **14**(1), 29-61 (2004)

Corbett, A., McLaughlin, M., Scarpinatto, C.: Modeling Student Knowledge: Cognitive Tutors in High School and College. User Modeling and User-Adapted Interaction **10**(2-3), 81-108 (2000)

Corbett, A.T., Anderson, J.R.: Student modeling and mastery learning in a computer-based programming tutor. In: Frasson, C., Gauthier, G., McCalla, G. (eds.) Second International Conference on Intelligent Tutoring Systems, ITS'92, Berlin, June 10-12, 1992 1992. Lecture Notes in Computer Science, pp. 413-420. Springer-Verlag, artc078 (k209a,b)

Corbett, A.T., Anderson, J.R.: Knowledge tracing: Modelling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction **4**(4), 253-278 (1995)

Corbett, A.T., Anderson, J.R., Carver, V.H., Brancolini, S.A.: Individual differences and predictive validity in student modeling. In: Ram, A., Eiselt, K. (eds.) the Sixteenth Annual Conference of the Cognitive Science Society, Edinburgh, 23-27 August 1993 1993a, pp. 457-464. Lawrence Erlbaum., artc146

Corbett, A.T., Anderson, J.R., O'Brien, A.T.: The predictive validity of student modeling in the ACT programming tutor. In: Brna, P., Ohlsson, S., Pain, H. (eds.) AI-ED'93, World Conference on Artificial Intelligence in Education, Charlottesville, Edinburgh, 23-27 August 1993 1993b, pp. 457-464. AACE, artc146

Cristea, A., Aroyo, L.: Adaptive authoring of adaptive educational hypermedia. In: De Bra, P., Brusilovsky, P., Conejo, R. (eds.) Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2002), Berlin, May 29-31, 2002 2002. Lecture Notes in Computer Science, pp. 122-132. Springer-Verlag

Davidovic, A., Warren, J., Trichina, E.: Learning benefits of structural example-based adaptive tutoring systems. IEEE Trans. Educ. **46**(2), 241-251 (2003)

De Bra, P.: Pros and Cons of Adaptive Hypermedia in Web-Based Education. Journal on CyberPsychology and Behavior **3**(1), 71-77 (2000)

De Bra, P., Aerts, A., Berden, B., de Lange, B., Rousseau, B., Santic, T., Smits, D., Stash, N.: AHA! The Adaptive Hypermedia Architecture. In: the Fourteenth ACM Conference on Hypertext and Hypermedia, Nottingham, UK 2003, pp. 81-84. ACM

De Bra, P., Ruiter, J.-P.: AHA! Adaptive hypermedia for all. In: Fowler, W., Hasebrook, J. (eds.) WebNet'2001, World Conference of the WWW and Internet, Orlando, FL, October 23-27, 2001 2001, pp. 262-268. AACE

De Bra, P., Smits, D., van der Sluijs, K., Cristea, A., Foss, J., Glahn, C., Steiner, C.M.: GRAPPLE: Learning Management Systems Meet Adaptive Learning Environments. In: Peña-Ayala, A. (ed.) Intelligent and Adaptive Educational Learning Systems: Achievements and Trends. (2013)

Eliot, C., Neiman, D., Lamar, M.: Medtec: A Web-based intelligent tutor for basic anatomy. In: Lobodzinski, S., Tomek, I. (eds.) WebNet'97, World Conference of the WWW, Internet and Intranet, Toronto, Canada, November 1-5, 1997 1997, pp. 161-165. AACE

Falakmasir, M.H., Pardos, Z.A., Gordon, G.J., Brusilovsky, P.: A Spectral Learning Approach to Knowledge Tracing. In: D'Mello, S.K., Calvo, R.A., Olney, A. (eds.) the 6th International Conference on Educational Data Mining (EDM 2013), Memphis, TN, USA, July 6-9, 2013 2013, pp. 28-34

Farzan, R., Brusilovsky, P.: AnnotatEd: A social navigation and annotation service for web-based educational resources. New Review in Hypermedia and Multimedia 14(1), 3-32 (2008)

Fogarty, J., Baker, R., Hudson, S.: Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction. In: Proceedings of Graphics Interface 2005. pp. 129-136. (2005)

Gena, C., Weibelzahl, S.: Usability Engineering for the Adaptive Web. In: Brusilovsky, P., Kobsa, A., Neidl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization, vol. 4321. Lecture Notes in Computer Science, pp. 720-762. Springer-Verlag, Berlin Heidelberg New York (2007)

Goldstein, I.P.: The Genetic graph: a representation for the evolution of procedural knowledge. Int. J. Man-Machine Studies 11(1), 51-77 (1979)

Goldstein, I.P., Carr, B.: The computer as coach: an athletic paradigm for intelligent education. In: 1977 annual ACM conference, Seatle, October 1977 1977, pp. 227-233, artc004

González-Brenes, J.P., Huang, Y., Brusilovsky, P.: General Features in Knowledge Tracing to Model Multiple Subskills, Temporal Item Response Theory, and Expert Knowledge. In: Stamper, J., Pardos, Z., Mavrikis, M., McLaren, B.M. (eds.) the 7th International Conference on Educational Data Mining (EDM 2014), London, UK, July 4-7, 2014 2014, pp. 84-91

Hatala, M., Gasevic, D., Siadaty, M., Jovanovic, J., Torniai, C.: Can Educators Develop Ontologies Using Ontology Extraction Tools: An End-User Study. In: Cress, U., Dimitrova, V., Specht, M. (eds.) 4th European Conference on Technology Enhanced Learning (ECTEL 2009), Nice, France, September 29- October 2, 2009 2009. Lecture Notes in Computer Science, pp. 127-139. Springer-Verlag

Heathcote, A., Brown, S., Mewhort, D.J.K.: The power law repealed: The case for an exponential law of practice. Psychonomic Bulletin & Review 7(2), 185-207 (2000)

Henze, N., Naceur, K., Nejdl, W., Wolpers, M.: Adaptive hyperbooks for constructivist teaching. Künstliche Intelligenz(4), 26-31 (1999)

Henze, N., Nejdl, W.: Student modeling for KBS Hyperbook system using Bayesian networks. In. University of Hannover, Hannover, Germany, (1999)

Hothi, J., Hall, W., Sly, T.: A study comparing the use of shaded text and adaptive navigation support in adaptive hypermedia. In: Brusilovsky, P., Stock, O., Strapparava, C. (eds.) Adaptive Hypermedia and Adaptive Web-based systems, Berlin, August 28-30, 2000 2000. Lecture Notes in Computer Science, pp. 335-342. Springer-Verlag

Hovland, C.I., Lumsdaine, A.A., Sheffield, F.D.: A baseline for measurement of percentage change. In: Hovland, C.I., Lumsdaine, A.A., Sheffield, F.D. (eds.) Experiments on mass communication. Wiley, (1949)

Hsiao, I.-H., Sosnovsky, S., Brusilovsky, P.: Adaptive Navigation Support for Parameterized Questions in Object-Oriented Programming. In: Cress, U., Dimitrova, V., Specht, M. (eds.) 4th European Conference on Technology Enhanced Learning (ECTEL 2009), Nice, France, September 29- October 2, 2009 2009. Lecture Notes in Computer Science, pp. 88-98. Springer-Verlag

Hsiao, I.-H., Sosnovsky, S., Brusilovsky, P.: Guiding students to the right questions: adaptive navigation support in an E-Learning system for Java programming. J. Comput. Assist. Learn. **26**(4), 270-283 (2010)

Jameson, A.: Modeling both the context and the user. Personal Technologies **5**(1), 29-33 (2001)

Karagiannidis, C., Sampson, D.G.: Layered evaluation of adaptive applications and servers. In: Brusilovsky, P., Stock, O., Strapparava, C. (eds.) Adaptive Hypermedia and Adaptive Web-based systens, Berlin, August 28-30, 2000 2000. Lecture Notes in Computer Science, pp. 343-346. Springer-Verlag

Kavcic, A.: Fuzzy User Modeling for Adaptation in Educational Hypermedia. IEEE Transactions on Systems, Man, and Cybernetics **34**(4), 439-449 (2004)

Keane, M., O'Brien, M., Smyth, B.: Are people biased in their use of search engines? Commun. ACM **51**(2), 49-52 (2008)

Knutov, E., De Bra, P., Pechenizkiy, M.: AH 12 years later: a comprehensive survey of adaptive hypermedia methods and techniques. New Rev. Hypermedia & Multimedia **15**(1), 5-38 (2009)

Koedinger, K., Mathan, S.: Distinguishing qualitatively different kinds of learning using log files and learning curves. In: Mostow, J., Tedesco, P. (eds.) TS 2004 Log Analysis Workshop, Maceio, Brazi 2004, pp. 39-46

Koedinger, K.R., Anderson, J.R., Hadley, W.H., Mark, M.A.: Intelligent tutoring goes to school in the big city. Int. J. AI in Education **8**, 30-43 (1997)

Lawless, S., Hederman, L., Wade, V.: Enhancing Access to Open Corpus Educational Content: Learning in the Wild. In: The 19th ACM Conference on Hypertext & Hypermedia, Pittsburgh, Pennsylvania, USA, June 19-21, 2008 2008, pp. 167-174

Martin, B., Koedinger, K.R., Mitrovic, A., Mathan, S.: On using learning curves to evaluate ITS. In: Looi, C.-K., McCalla, G., Bredeweg, B., Breuker, J. (eds.) 12th International Conference on Artificial Intelligence in Education, AIED'2005, Amsterdam, July 18-22, 2005 2005, pp. 419-426. IOS Press

Martin, B., Mitrovic, A., Koedinger, K., Mathan, S.: Evaluating and Improving Adaptive Educational Systems with Learning Curves. User Modeling and User-Adapted Interaction **21**(3), 249-283 (2011).

Mitrovic, A.: An Intelligent SQL Tutor on the Web. Int. J. AI in Education **13**(2-4), 173-197 (2003)

Mitrovic, A., Ohlsson, S.: Evaluation of a constraint-based tutor for a database language. Int. J. AI in Education **10**(304), 238-256 (1999)

Murre, J., Chessa, A.: Power laws from individual differences in learning and forgetting: mathematical analyses. Psychonomic Bulletin & Review **18**(3), 592-597 (2011).

Newell, A., Rosenbloom, P.S.: Mechanisms of skill acquisition and the law of practice. In: Anderson, J.R. (ed.) Cognitive skills and their acquisition. pp. 1-51. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ (1981)

Oberlander, J., O'Donell, M., Mellish, C., Knott, A.: Conversation in the museum: experiments in dynamic hypermedia with the intelligent labeling explorer. The New Review of Multimedia and Hypermedia **4**, 11-32 (1998)

Ohlsson, S.: Constraint-based student modeling. J. Artif. Intellig. in Educ. **3**(4), 429-447 (1992)

Olston, C., Chi, E.H.: ScentTrails: Integrating browsing and searching on the Web. ACM Trans. Comput.-Hum. Interact. **10**(3), 177-197 (2003)

Papanikolaou, K.A., Grigoriadou, M., Kornilakis, H., Magoulas, G.D.: Personalising the interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. User Mod. & User-Adapt. Interact. **13**(3), 213-267 (2003)

Paramythis, A., Weibelzahl, S.: A Decomposition Model for the Layered Evaluation of Interactive Adaptive Systems. In: Ardissono, L., Brna, P., Mitrovic, A. (eds.) 10th International User Modeling Conference, Edinburgh, UK, July 24-29, 2005 2005. Lecture Notes in Artificial Intelligence, pp. 438-442. Springer Verlag

Pardos, Z.A., Heffernan, N.: KT-IDEM: introducing item difficulty to the knowledge tracing model. In: Konstan, J., Conejo, R., Marzo, J., Oliver, N. (eds.) 19th International Conference on User Modeling, Adaptation, and Personalization, UMAP 2011, Girona, Spain, July 11-15, 2011 2011, pp. 243-254. Springer-Verlag

Pavlik, P., Cen, H., Koedinger, K.: Performance Factors Analysis - A New Alternative to Knowledge Tracing. In: Proceedings of Artificial Intelligence in Education 2009. pp. 531–538. (2009)

Pirolli, P., Kairam, S.: A knowledge-tracing model of learning from a social tagging system. User Modeling and User-Adapted Interaction **23**(2-3), 139-168 (2013).

Ritter, S.: PAT Online: A Model-tracing tutor on the World-wide Web. In: Brusilovsky, P., Nakabayashi, K., Ritter, S. (eds.) Workshop "Intelligent Educational Systems on the World Wide Web" at AI-ED'97, 8th World Conference on Artificial Intelligence in Education, Kobe, Japan, 18 August 1997 1997, pp. 11-17, also available at http://www.contrib.andrew.cmu.edu/~plb/AIED97_workshop/Ritter/Ritter.html. ISIR

Schneider-Hufschmidt, M., Kühme, T., Malinowski, U. (eds.): Adaptive user interfaces: Principles and practice. North-Holland, Amsterdam (1993)

Self, J.: Student models in computer-aided instruction. Int. J. Man-Machine Studies **6**, 261-276 (1974)

Self, J.: Bypassing the intractable problem of student modelling. In: Frasson, C., Gauthier, G. (eds.) Intelligent Tutoring Systems: At the crossroads of artificial intelligence and education. pp. 107-123. Ablex Publishing, Norwood (1990)

Shute, V.J.: SMART: Student modeling approach for responce tutoring. User Mod. & User-Adapt. Interact. **5**(1), 1-44 (1995)

Smith, A.S.G., Blandford, A.: MLTutor: An application of machine learning algorithms for an adaptive Web-based information system. Int. J. AI in Education **13**(2-4), 235-261 (2003)

Sosnovsky, S., Brusilovsky, P.: Layered Evaluation of Topic-Based Adaptation to Student Knowledge. In: Fourth Workshop on the Evaluation of Adaptive Systems at 10th International User Modeling Conference, UM 2005, July 26, 2005 2005, pp. 47-56

Sosnovsky, S., Brusilovsky, P., Hsiao, I.-H.: Adaptation "in the Wild": Ontology-based Personalization of Open-Corpus Learning Material. In: 7th European Conference on Technology Enhanced Learning (EC-TEL 2012), Saarbrücken, Germany 2012. Lecture Notes in Computer Science, pp. 425-431

Sosnovsky, S., Brusilovsky, P., Lee, D.H., Zadorozhny, V., Zhou, X.: Re-assessing the Value of Adaptive Navigation Support in E-Learning. In: Nejdl, W., Kay, J., Pu, P., Herder, E. (eds.) 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2008), Hannover, Germany, July 29-August 1, 2008 2008. Lecture Notes in Computer Science, pp. 193-203. Springer Verlag

Sosnovsky, S., Brusilovsky, P., Yudelson, M.: Supporting Adaptive Hypermedia Authors with Automated Content Indexing. In: Second International Workshop on Authoring of Adaptive and Adaptable Educational Hypermedia at the Third International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH'2004), Eindhoven, the Netherlands 2004

Specht, M., Kravcik, M., Pesin, L., Klemke, R.: Learner's Lounge: Information Brokering for the Adaptive Learning Environment. In: Driscoll, M., Reeves, T.C. (eds.) World Conference on E-Learning, E-Learn 2002, Montreal, Canada, October 15-19, 2002 2002, pp. 2510-2512. AACE

Specht, M., Oppermann, R.: ACE - Adaptive Courseware Environment. The New Review of Hypermedia and Multimedia **4**, 141-161 (1998)

Specht, M., Weber, G., Heitmeyer, S., Schöch, V.: AST: Adaptive WWW-Courseware for Statistics. In: Brusilovsky, P., Fink, J., Kay, J. (eds.) Workshop "Adaptive Systems and User Modeling on the World Wide Web" at 6th International Conference on User Modeling, UM97, Chia Laguna, Sardinia, Italy, June 2, 1997 1997, pp. 91-95, also available at http://www.contrib.andrew.cmu.edu/~plb/UM97_workshop/Specht.html

Srivastava, J., Cooley, R., Deshpande, M., Tan, P.: Web usage mining: discovery and applications of usage patterns from Web data. SIGKDD Explor. Newsl. **1**(2), 12-23 (2000)

Stern, M.K., Woolf, B.P.: Curriculum sequencing in a Web-based tutor. In: Goettl, B.P., Halff, H.M., Redfield, C.L., Shute, V.J. (eds.) 4th International Conference, ITS-98, Berlin, August 16-19, 1998 1998. Lecture Notes in Computer Science, pp. 574-583. Springer-Verlag

Triantafillou, E., Pomportis, A., Demetriadis, S., Georgiadou, E.: The value of adaptivity based on cognitive style: an empirical study. Brit J Educ Technol **35**(1), 95–106 (2004)

VanLehn, K.: Student models. In: Polson, M.C., Richardson, J.J. (eds.) Foundations of intelligent tutoring systems. pp. 55-78. Lawrence Erlbaum Associates, Hillsdale (1988)

Vassileva, J.: DCG + GTE: Dynamic Courseware Generation with Teaching Expertise. Instr. Science **26**(3/4), 317-332 (1998)

Virvou, M., Moundridou, M.: Student and instructor models: Two kinds of user models and their interaction in an ITS authoring tools. In: Bauer, M., Gmytrasiewicz, P.J., Vassileva, J. (eds.) 8th International Conference on User Modeling, UM 2001, Berlin, July 13-17, 2001 2001. Lecture Notes on Artificial Intelligence, pp. 158-167. Springer-Verlag

Wang, J.Z., Taylor, W.: Concept Forest: A New Ontology-Assisted Text Document Similarity Measurement Method. In: Lin, T.Y., Haas, L., Kacprzyk, J., Motwani, R., Broder, A., Ho, H. (eds.) the 2007 international conference on Web Intelligence, WI '07, Silicon Valey, CA, USA, November 2-5, 2007 2007, pp. 395-401. IEEE

Weber, G., Kuhl, H.-C., Weibelzahl, S.: Developing adaptive internet based courses with the authoring system NetCoach. In: Bra, P.D., Brusilovsky, P., Kobsa, A. (eds.) Third workshop on Adaptive Hypertext and Hypermedia, Sonthofen, Germany, July 14, 2001 2001, pp. 35-48, also available at http://wwwis.win.tue.nl/ah2001/papers/GWeber-UM2001.pdf

Weber, G., Kuhl, H.-C., Weibelzahl, S.: Developing adaptive internet based courses with the authoring system NetCoach. In: Reich, S., Tzagarakis, M.M., De Bra, P.M.E. (eds.) Hypermedia: Openness, Structural Awareness, and aptivity. pp. 226-238. Springer-Verlag, Berlin (2002)

Yudelson, M.: Providing service-based personalization in an adaptive hypermedia. University of Pittsburgh (2010)

Yudelson, M., Koedinger, K., Gordon, G.: Individualized Bayesian Knowledge Tracing Models. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) Proceedings of Artificial Intelligence in Education 2013. pp. 171-180. Springer, Berlin/Heidelberg, Germany (2013)