

Feedback Personalization as Prerequisite for Assessing Higher-order Thinking Skills

Christian Saul, [christian.saul@idmt.fraunhofer.de],
Fraunhofer Institute for Digital Media Technology, Germany
Heinz-Dietrich Wuttke, [dieter.wuttke@tu-ilmenau.de],
Faculty Informatics and Automation, Germany

Topics of the Paper

- Introduction
- Feedback Research
- Adaptive Assessment
 - Adaptive Testing
 - Adaptive Questions
 - Comparison of Adaptive Assessment Systems towards Feedback
- Thinking Skills
 - Assessing Thinking Skills
 - Comparison of Adaptive Assessment Systems towards the Assessment of Higher-order Thinking Skills
- Discussion
- Conclusions and Future Work

Introduction

Nowadays, personalization is increasingly becoming a crucial factor in many areas of life including education, health care and television. Almost every service is designed to accommodate preferences and expectations that are usually different between individuals. It is believed that personalization in education raises motivation and interests of students, which are critical success factors in the learning process. Personalized support for students becomes even more important, when learning takes place in open and dynamic learning and information network environments. In this context, a manually performed personalization is too time consuming and thus the use of information technologies appears to be a necessity in personalized education. Over time, several Educational Adaptive Hypermedia Systems (EAHS) namely *Interbook* (Brusilovsky et al., 1998), *AHA!* (De Bra & Calvi, 1998) or *APeLS* (Conlan, 2005) were developed, which aimed at addressing personalization issues in learning context. EAHS build a model of the goals, preferences and knowledge of each student and use this model throughout the interaction with the student in order to adapt the system as well as the learning content to the needs of that student.

Although personalization in educational settings is well advanced, it is still neglected in assessments. Assessment is defined as any systematic method of obtaining evidence by posing questions to draw inferences about the knowledge, skills, attitudes and other characteristics of people for a specific purpose (Shepherd & Godwin, 2004). Stand-alone applications that are designed to be delivered across the web for assessing students' learning are called online-assessments. Online-assessments enable the assessors observing and automatically evaluating students' progress. This results in reduced economical costs by cost savings in room and staff necessary for supporting and correcting, time savings in correcting the results as well as material savings through digitalization. Furthermore, online-assessments provide improved reliability, because automated marking is much more reliable than human marking and enable enhanced question styles, which incorporate interactivity and multimedia. The integration of pictures, sounds and videos in online-assessment improves the clearness of questions and tests by the use of interactive scenarios and simulations. From the students' point of view, online-assessments help to learn by providing instant and detailed feedback, which serve as motivation and learning aid. Additionally, online-assessments offer students increased flexibility with respect to location and timing. But online-assessments have reached their limits when it comes to considering individual and social aspects. A fully automated process and the loss of personal contact and support can be frustrating for the students and thus can cause the feeling of getting lost in the masses.

Although several online-assessment systems indicate some aspects of personalization (Brusilovsky et al., 2004; Cheniti-Belcadhi et al., 2008; Conejo et al., 2004), personalized assessment goes a step further. Issues such as subjects of the tasks, levels of difficulty and feedback should be adapted to the students'

individual context, prior knowledge and preferences.

With respect to feedback adaptation, only a few studies were investigated: Lütticke (2004) has experimentally demonstrated the effectiveness of feedback adaptation in a problem-solving task. He adapted the content of feedback to the students' individual errors, knowledge, preferences in support and progress in solving the problem. The experiments showed that 80% of the students favour feedback adaptation and most of them wish to have more adaptive feedback. Chuang and O'Neil (2006) also performed a study to investigate various types of feedback. More than 120 students were asked to search a web environment of information and to improve a knowledge map. The study clearly showed that students, which got adaptive feedback, performed better than students, which got no adaptive feedback. To summing up, the results of the experiments seem to suggest that the perspectives of feedback adaptation for web-based systems are promising, in particular for online-assessment systems.

These were the reasons why the authors have decided to investigate adaptive (online-) assessment systems providing personalized feedback. The focus in this paper is to analyze the incorporation of feedback personalization in adaptive assessment systems and possibly to point out potential areas for improvement in this respect.

The remainder of the paper is organized as follows: The second chapter gives an insight in feedback research and proposes a 3-dimensional feedback classification. The third chapter describes four existing AASs (SIETTE, PASS, CosyQTI and iAdaptTest) and provides an analysis of these systems according to the previous defined feedback classification. The fourth chapter investigates thinking skills as well as how these systems address these skills. The fifth chapter discusses these findings and concluding remarks and references complete the paper.

It is important to note that the term student in this paper means everybody aiming at acquiring, absorbing and exchanging knowledge, whereas learning is to be understood likewise. Hence, the explanations and conclusions in this paper are not limited to typical teacher-student relationships, but also applicable to any kind of knowledge provider and knowledge consumer.

Feedback Research

Feedback plays an important role not only in education but also in various fields of science including psychology, biology and economics. Generally, feedback is studied within human-computer interaction based upon two problems: how to organize feedback to the user and how to predict and process feedback from the user (Vasilyeva et al., 2007). The focus of this paper lies on the former problem. In education, the main aim of feedback can be defined as informing and motivating the student to increase their effort and attention. Further, feedback is fundamental for information systems design, because it constitutes an important part of how users experience a system (Sharp et al., 2007). Adequate feedback increases the users feeling of being in control (Benyon et al., 2005). In the case of personalized support for students, feedback would indicate whether the student actually is learning and keeping the right track. As such, it can make the difference between using and not using a system.

Kulhavy and Wagner (1993) introduced the concept of a feedback-triad, which included three definitions of feedback: feedback as a motivator for increasing response rate and/or accuracy, feedback reinforcing a message that would automatically connect responses to prior stimuli and feedback providing information that students could use to validate or change a previous response. The study clearly demonstrates the nature of the feedback problem. The users stated that feedback should function and be analyzable on several levels: as a motivator, provider of information and reinforcement. Black and Wiliam (1998) distinguished four elements of a feedback system: data on the actual level of some measureable attribute (students' answer to a question), data on the reference level of that attribute (the correct answer), a mechanism for comparing the two levels and generating information about the gap between the two levels and a mechanism by which the information can be used to alter the gap (as it presents help to a student in the case of an incorrect answer). Iahad et al. (2004) defined feedback as rich, if it provides feedback through automatic grading, if it provides correct answers and if it refers the students to the learning content, which explains the correct answers.

In the literature, different types of feedback classifications have been presented. A short excerpt is presented in the following: According to Kulhavy and Stock (1989), effective feedback provides the student with two types of information: *verification* and *elaboration*. Verification is the simple judgment of whether an answer is correct or incorrect, while elaboration is the informational component providing relevant cues to guide the student toward a correct answer. Elaborative feedback can be used in form of hints and represents a kind of stimuli towards the correct answer. Feedback elaboration is typically *informational*, *topic-specific* or *response-specific*. Moreover, feedback can take on many forms depending on the levels of verification and elaboration incorporated. According to Mason and Bruning (2001), feedback can be distinguished into: *no-feedback*, *knowledge-of-response*, *answer-until-correct*, *knowledge-of-correct-response*, *topic-contingent* and *response-contingent*. No-feedback simply provides students with the performance score with no reference to individual test items. This minimal level of feedback contains neither verification nor elaboration, but simply states the students' number or proportion of correct responses. Knowledge-of-response tells students whether their answers are correct or incorrect. While this

type of feedback is essential for verification purposes, it does not provide any information that would extend the students' knowledge or provide additional insight into possible errors in understanding. Answer-until-correct feedback provides verification but no elaboration and requires the student to remain on the same test item until the correct answer is given. Knowledge-of-correct-response feedback provides individual question verification and supplies students with the correct answer, but does not offer any elaborative information. Topic-contingent feedback provides item verification and general elaborative information concerning the target topic. Response-contingent feedback gives response-specific feedback that explains why the incorrect answer was wrong and why the correct answer is correct. According to Dempsey and Wager (1988), feedback can also be classified into *immediate* and *delayed*. Immediate feedback is presented to the student immediately after the answer is given. In contrast, delayed feedback is presented after a specified delay interval during testing. In addition, feedback can be differentiated according to the form of presentation used: *textual*, *graphical*, *auditory* and *animated* or a combination of these (Sharp et al., 2007). Textual feedback like 'ok' or 'well done' in case of correct answer and 'no' or 'try again' in the opposite case is the most commonly used form of feedback presentation. Graphical feedback is often used in computer games and illustrates the completed levels or progress. Animated feedback is typical used in multimedia systems as well as computer games. For a more comprehensive overview of feedback classifications, reference is made to Mory (2004) and Vasilyeva et al. (2007).

Analyzing the different feedback classifications, feedback can be categorized into three dimensions: *response*, *occurrence* and *presentation*. These facts are graphical represented in Figure 1.

Feedback plays a central role in the assessment process, because it provides information about the current areas of strength and weakness of the particular students. Feedback can be regarded as the so called speaking tube of the question and test evaluation and thus able to communicate the result of the assessment to the students as well as other information, which may contain reasons for incorrect answers, hints or advices for continuing the assessment. The next chapter investigates how four established AASs deal with feedback and especially how comprehensive they cover the three dimensions of feedback.

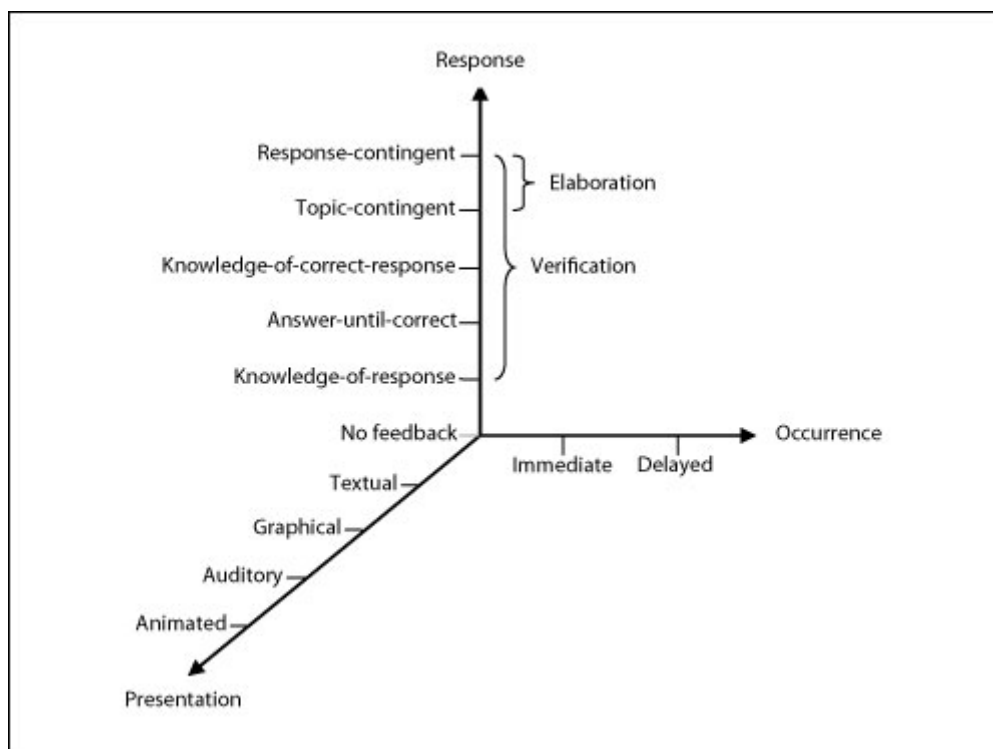


Figure 1 Dimensions of Feedback

Adaptive Assessment

There is a demand towards personalization in online-assessment to take care of the individual needs and avoid treating all students in the same manner. An AAS poses one way to realize personalization in online-assessments. AASs and technologies are used to test students at their current knowledge level and change their behaviour and structure depending on the students' previous responses, individual context, prior knowledge and preferences. There are two types of adaptive techniques that can be applied in AASs namely adaptive testing (Wainer et al., 2000; Linden & Glas, 2000) and adaptive questions (Pitkow & Recker, 1995).

Adaptive Testing

The adaptive testing technique involves a computer-administered test in which the selection and

presentation of each question and the decision to stop the process are dynamically adapted to the student's performance in the test. The technique uses a statistical model, mostly the Item Response Theory (Hambleton et al., 1991), to estimate the probability of a correct answer to a particular question and to select an appropriate question accordingly. Appropriate questions are selected from a pool of questions so that their difficulty matches the students' estimated level of knowledge. The questions that provide most amount of information about the current knowledge level of the student are usually those with difficulty similar to the students' knowledge level (Bloom et al., 1956). An advantage of adaptive testing is that questions, which are too difficult or too easy, are removed. Thus, the technique ensures that the student only sees questions that are very close to his or her level of knowledge. However, the technique only supports multiple-choice or true-false questions. It is not designed for advanced question types. Several approaches exploit the technique of adaptive testing such as SIETTE (Conejo et al., 2004) and PASS (Gouli et al., 2002).

SIETTE is one of the first web-based tools, which assists authors of questions and tests in the assessment process and adapts to the students' current level of knowledge. The system uses Java Applets for authoring and presenting adaptive tests. In SIETTE, the selection of questions is based on a function that estimates the probability of a correct answer to a particular question, which leads to an estimation of the students' level of knowledge. The question with the highest probability will be posed. Although SIETTE infers students' knowledge level through adaptive testing and presents questions to the student adapted to the current level of knowledge, the system has some disadvantage in terms of estimating students' knowledge level separated to the particular topics in a test. It mainly uses multiple-choice questions and provides only insufficient support in terms of feedback and help.

PASS (*Personalized ASSESSment*) is a web-based assessment module, which can be integrated into an adaptive educational hypermedia system to provide personalized assessment. The system estimates students' performance through multiple assessment options (pre-test, self-assessment and summative assessment) tailored to students' responses. The system enables the educators to define assessment specifications and to have a detailed overview of the students' performance and progress. Advantageous of PASS is the consideration of the students' navigational behaviour, the re-estimation of the difficulty level of each question at any time it is posed as well as the consideration of the importance of each educational material page. However, the feedback provided to the students is not adapted to their performance and thus lacks personalization.

Adaptive Questions

The adaptive questions technique defines a dynamic sequence of questions depending on students' responses. The technique defines rules, which allow selecting questions dynamically. The defined rules are linked, for example, to the response of the student and an overlay student model, which represents student knowledge of different concepts and topics. Based on these rules and the last response of the student, appropriate questions can dynamically be selected at runtime. The technique of adaptive questions offers more flexibility than the technique of adaptive testing, because authors of tests are given the flexibility to express their didactical philosophy and methods through the creation of appropriate rules. Several approaches exploit the technique of adaptive questions such as CosyQTI (Lalos et al., 2005) and iAdaptTest (Lazarinis et al., 2009).

CosyQTI is a web-based tool for authoring and presenting adaptive assessments based on IMS QTI (2006), IMS LIP (2005) and IEEE LTSC PAPI (2001) learning standards. The system consists of a *student model*, a *domain model* and a *rule model*. The *student model* contains information such as the goals, preferences, qualifications, knowledge estimations and usage data of each student. The *domain model* follows the IEEE/ACM vocabulary structure and allows educators of various disciplines utilizing the system. Adaptation decisions are set by the educators during the authoring phase by defining *IF <condition> THEN <action>* rules, which are contained in the *rule model*. Moreover, CosyQTI allows students to access parts of their profile and to raise the awareness of their current knowledge, strengths and weaknesses. The advantages of CosyQTI are the conformance to different established standards and specifications, which make the system interoperable with other standard-compliant learning tools and systems. Moreover, the open information policy leads to enhanced learning, but there are still some problems with authoring and selecting questions. Regarding the authoring of questions, the limited rule system and the few question types restrict the incorporation of didactic philosophy and methods. Besides, the use of feedback in the assessment process is rather limited. In terms of question selecting, CosyQTI is relatively weak in estimating and representing students' current knowledge level.

iAdaptTest is a desktop-based modularized adaptive testing tool conforming to the IMS QTI (2006), the IMS LIP (2005) and XML Topic Maps (2001) in order to improve the reusability and interoperability of the data. The data are stored in distinct files and can independently be shared across different learning tools and systems. Although iAdaptTest is entirely based on established standards and specifications, the system has still some problems. The first one is that it has been implemented as a Microsoft Windows application, which means that it can only be used on Microsoft Windows operation systems. In addition, iAdaptTest provides only a few question types and the implemented feedback and help is rather simple and does not enable personalized support.

Comparison of Adaptive Assessment Systems towards

Feedback

The comparisons between the above mentioned AASs according to the different feedback classifications are provided in Table 1. The table above shows that each of the AAS provides possibilities to incorporate feedback in the assessment process. However, the use of feedback techniques is limited. According to type of information (response), almost all systems are limited to the use of knowledge-of-correct-response feedback. Thus, the respective systems only provide verifying information in form of correct responses. This type of feedback provides no elaborative information, for example, the part of the course, in which the subject of the question is described. According to the way of presentation, all systems use the textual way of presenting feedback. The authors of the systems are restricted to use several words like “ok”, “well done” or “not correct, try again” as the form of presenting feedback. According to the time of occurrence, all systems are restricted to immediate feedback. This means that feedback to the student is given immediately after answering and not delayed during testing.

As a result, SIETTE, PASS, CosyQTI and iAdaptTest provide possibilities to incorporate feedback in the assessment process, but they only use a limited set of feedback techniques (see Table 1) and do not take into account any students' individual characteristics or needs. This results in not exploiting the potential of personalization that feedback actually has. In order to determine the reasons for that, the next chapter will investigate thinking skills as well as how SIETTE, PASS, CosyQTI and iAdaptTest address these skills.

Table 1 Comparison of Adaptive Assessment Systems towards Feedback

Dimension		SIETTE	PASS	CosyQTI	iAdaptTest
Response	Response-contingent				
	Topic-contingent				
	Knowledge-of-correct-response	X	X	X	X
	Answer-until-correct				
	Knowledge-of-response				
Presentation	Textual	X	X	X	X
	Graphical				
	Animated				
	Auditory				
Occurrence	Immediate	X	X	X	X
	Delayed				

Thinking Skills

The term thinking skills refers to the human capacity to think in conscious ways to achieve certain purposes. Such processes include remembering, questioning, forming concepts, planning, reasoning, imagining, solving problems, making decisions and judgments as well as translating thoughts into words (Fisher, 2006). Thinking skills were conceptualized in a number of ways and at present there is little consensus with regard to the actual term. But, it is generally agreed that thinking skills can roughly be divided into lower-order (LOTS) and higher-order thinking skill (HOTS). HOTS are grounded in LOTS and linked to prior knowledge (King et al., 1998). HOTS include critical thinking, problem solving, decision making and creative thinking (Lewis & Smith, 1993). These skills are activated when students encounter unfamiliar problems, uncertainties, questions or dilemmas. Successful applications of these skills result in explanations, decisions and performances that are valid within the context of available knowledge and experience and promote continued growth in higher-order thinking as well as other intellectual skills.

In this paper, the efforts undertaken by Benjamin Bloom were used to differentiate thinking skills. In the 50s of the last century, he led a team of educational psychologists trying to analyze and classify the varied domains of human learning (cognitive, affective and psychomotor). The efforts resulted in a series of taxonomies in each domain, known today as Bloom's taxonomies (Bloom et al., 1956). The cognitive

domain involves knowledge and the development of intellectual skills. In this domain, Bloom et al. distinguish between six different levels namely *knowledge*, *comprehension*, *application*, *analysis*, *synthesis* and *evaluation*. The first three levels are referred to as LOTS and the last three levels are referred to as HOTS (King et al., 1998). More than 50 years later, Bloom's taxonomies of the cognitive domain were revised by Anderson and Krathwohl (Anderson et al., 2001). Differences are the rewording of the levels from nouns to verbs, the renaming of some of the components and the repositioning of the last two categories (see Table 2).

Table 2 Taxonomies of the Cognitive Domain

Bloom (1956)	Anderson and Krathwohl (2001)
Knowledge	Remember
Comprehension	Understand
Application	Apply
Analysis	Analyze
Synthesis	Evaluate
Evaluation	Create

The lowest, so called *remembering* level requires the students to recall and recognize terms and their place in a particular domain. The *understanding* level requires the students to inherit information from these terms by interpreting, summarizing or inferring. The *applying* level requires the students to use a learned topic in an appropriate situation. The *analyzing* level requires the students to separate the parts of a whole and to understand the relationships in between. The *evaluation* level requires the students to make judgments based on criteria and standards through checking and critiquing and the *creation* level requires the students to combine parts to create a new whole, where that whole is not apparent before creation. But, the major differences are the addition of how the taxonomy intersects and acts upon different types and levels of knowledge, namely *factual*, *conceptual*, *procedural* and *meta-cognitive* (see Table 4). *Factual knowledge* is knowledge that is basic to specific disciplines. It encompasses essential facts, terminology or details students must know order to understand a discipline or solve a problem. *Conceptual knowledge* is knowledge about the interrelationships among the basic elements within a larger structure that enable them to function together. *Procedural knowledge* is knowledge that helps students to do something. It consists of criteria for using skills, algorithms, techniques and methods. *Meta-cognitive knowledge* is knowledge of cognition in general as well as awareness of one's own cognition.

Assessing Thinking Skills

Assessment is regarded as very useful for measuring LOTS such as recall and interpreting of knowledge, but seen as insufficient for assessing HOTS such as the ability to apply knowledge in new situations or to evaluate and synthesize information. But, this need not be the case. Sugrue (1995) identified three response formats for measuring HOTS namely *selection*, *generation* and *explanation*. *Selection* means using simple question types such as multiple-choice and matching for identifying the most plausible assumption or the most reasonable inference. *Generation* means using advanced question types, which let students more creativity in answering, such as free-text answers, essays and interactive and simulative tools for measuring HOTS and *explanation* means giving reasons for selection or generation of a response. This is often realized by asking for an additionally written justification of the answer.

In addition to the even explained response formats, it is crucial that the students have sufficient prior knowledge, because it serves as basis for using their HOTS in answering questions or performing tasks. For that reason, assessments that address HOTS should adapt for diverse student needs. They should support at the beginning and then gradually turning over responsibility to the students to operate on their own (Kozloff & Wilmington, 2002). This limited temporary support helps students develop HOTS.

As it is generally agreed that assessment systems and in particular AASs are able to assess LOTS (for example, recall and interpreting of knowledge), in the following, special attention is laid on the assessment of HOTS by SIETTE, PASS, CosyQTI and iAdaptTest.

Comparison of Adaptive Assessment Systems towards the Assessment of Higher-order Thinking Skills

As mentioned earlier, there are three response formats for measuring HOTS namely selection, generation

and explanation. As the presence of these formats indicate the potential for addressing HOTS during the assessment process, the comparison was focused on these criteria. The results of the comparison are provided in Table 3. The table shows that that each of the AAS is limited to the selection response format. That means that they only provide simple question types. SIETTE and PASS only admit traditional multiple-choice questions without any written justification (explanation). This is due to the fact that they use the technique of adaptive testing, which only supports multiple-choice or true-false questions and is not designed for advanced question types (generation). CosyQTI allows creating true-false, multiple-choice, single-, multiple and ordered response as well as image hot spot questions. The question types provided by iAdaptTest are similar to CosyQTI, namely true-false, single-, and multiple-choice, gap match and association. As CosyQTI and iAdaptTest follow the adaptive questions technique, they are less restricted in providing advanced question types compared to SIETTE and PASS. However, they do not allow the creativity in answering as required by the generation response format. Additionally, both systems do not include any form of question justification necessary for the explanation response format.

Table 3 Comparison of SIETTE, PASS, CosyQTI and iAdaptTest towards the Assessment of Higher-order Thinking Skills

Response Format	SIETTE	PASS	CosyQTI	iAdaptTest
Selection	X	X	X	X
Generation				
Explanation				

Summarized, this means that although all analyzed AASs can be used for measuring some specific HOTS such as deduction, inference and prediction, they are inappropriate for measuring skills on the evaluation and creation level. With respect to the taxonomies presented above, the potential of these AASs for assessing thinking skills is presented in Table 4. The table illustrates that SIETTE, PASS, CosyQTI and iAdaptTest have the potential for assessing thinking skills on the remembering, understanding, applying and limited on the analyzing level in all knowledge dimensions.

Table 4 Taxonomy Matrix of SIETTE, PASS, CosyQTI and iAdaptTest (adapted from Anderson et al.)

		Cognitive Process Dimension					
		LOTS			HOTS		
		Remember	Understand	Apply	Analyze	Evaluate	Create
Knowledge Dimension	Factual	X	X	X	(X)		
	Conceptual	X	X	X	(X)		
	Procedural	X	X	X	(X)		
	Meta-cognitive	X	X	X	(X)		

Discussion

Each of the AASs investigated, presented and compared in this paper estimates the knowledge level of each student and based upon the system selects appropriate questions using different approaches and techniques. There are many systems using the number of questions answered correctly and the difficulty level of answered questions, such as SIETTE and PASS. By contrast, other systems such as CosyQTI and iAdaptTest define rules, which allow selecting questions dynamically.

Although the majority of these systems tailor the selection of question within the assessment process to the knowledge level of each student, personalization with regard to feedback is almost entirely disregarded. The comparison of these systems towards feedback substantiates this statement. All systems are restricted to provide knowledge-of-correct-response feedback. They usually provide feedback in forms of simply telling if the answer is correct, not correct or partially correct as well as giving the correct answer. Although knowledge-of-correct-response feedback not only provides feedback regarding whether the answer was received or not (knowledge-of-response), but also whether the answer was correct or not, it does not provide additional information. But elaborative feedback is essential when striving for implementing feedback that is adapted to the individual students' context. Elaborative feedback could be realized, for

example, through a virtual coach, which appears at the end of a question block and presents a summary of completed questions as well as hints or advices for continuing the assessment. This intensifies the dynamic behaviour of the system resulting in the feeling of the students to communicate with another actor. This can be compared with oral examinations, in which the assessor provides additional information that is important for completing the task, but does not immediately offer the correct solution. This fact can be verified by referring to Kulhavy and Stock (1989). They demonstrated significant improvements in learning using elaborative feedback. Concerning the timing of the feedback, some researchers argue that immediate feedback is needed to maintain the students' attention and motivation (Corbett & Anderson, 2001), while in earlier research others have shown that delayed feedback can contribute to better retention and transfer of skills (Kulhavy & Anderson, 1972). Osborne and Winkley (2006) also stated that a good online-assessment system provides the student with immediate and relevant feedback at the point of error in order to take advantage of the lessons learned. As analyzed above, all investigated AASs are limited to provide immediate feedback. They present the feedback to the student immediately after the answer is given. Although immediate feedback seems to be more effective than delayed feedback, they could benefit from each other if immediate verification feedback is combined with delayed elaborative feedback. This enables students to have immediate knowledge about the correctness of their response, but they still have time to think about errors before elaborative information is given. With respect to the way of presenting feedback, all investigated AASs make use of textual feedback and do not provide possibilities to integrate other forms of feedback presentation like graphics, animations, videos or sounds. But with respect to a personalization of feedback, these forms are of particular importance. Czerwinski and Larson (2003) argued that these forms of feedback increase the attention and can motivate the students. It is also important to note that the provision of feedback must be carefully provided in order to prevent unintended influence of the student. The feedback should not affect the students in such a way that they are no longer able to answer questions independently, but instead make their decisions according to the provided information.

As shown, feedback has an enormous potential in realizing personalization in assessments. But, what are the reasons of SIETTE, PASS, CosyQTI and iAdaptTest to not making use of them. On this account, thinking skills their addressing by these AASs were investigated. As a result, SIETTE, PASS, CosyQTI and iAdaptTest are able to assess LOTS, but they are inappropriate for measuring skills on the evaluation and creation level (HOTS). But, learning in the twenty-first century is about integrating and using knowledge and not just about acquiring facts and procedures (Fadel et al., 2007). For example, in engineering education, the students should be able to develop new technical systems. For that, they have to combine parts to create a new whole and to evaluate the results appraisingly (Wuttke et al., 2008). Furthermore, HOTS are essential for success not only in learning, but also in life (Fisher, 2006). Due to that fact, assessment systems and in particular AASs need to evaluate not just the students' factual knowledge (LOTS), but also their problem-solving and reasoning strategies (HOTS), which are currently left to oral examinations or project work. In addition to use the explained response formats (selection, generation and explanation) for addressing HOTS, it is crucial that assessments adapt for diverse student needs. This limited temporary support helps students develop HOTS. As SIETTE, PASS, CosyQTI and iAdaptTest do not really address HOTS (see Table 4), it is not surprising that they do not exploit the potential of personalization that feedback actually has (see Table 1). Finally, it can be stated that when striving for the assessment of HOTS of students, personalized support and in detail personalized feedback is essential.

Conclusions and Future Work

The objective of this paper was to analyze the incorporation of feedback personalization in AASs (SIETTE, PASS, CosyQTI and iAdaptTest) and possibly to point out potential areas for improvement in this respect. The analysis was caused by an understanding of the need of assessment adapted to the students' individual context, prior knowledge and preferences. Taking into account such criteria in order to personalize the assessment may result in more valid assessments and in particular in more objective assessment findings. Although these systems adapt the assessment process of each student resulting in presenting different questions they still enable a better comparability between different individuals, because each individual would be more correctly assessed. Moreover, they reveal the current areas of strength and weakness of the students more precisely.

The results of the analysis pointed out that SIETTE, PASS, CosyQTI and iAdaptTest provide possibilities to test students at their current knowledge level and change the systems' behavior and structure depending on the students' responses and detected abilities. But as shown, personalization of feedback is still insufficiently implemented or even not addressed in these systems. Reasons for that could be found in analyzing the thinking skills assessed. As shown, SIETTE, PASS, CosyQTI and iAdaptTest only address LOTS and are not appropriate for assessing HOTS. But, as learning in the twenty-first century is about integrating and using knowledge and not just about acquiring facts and procedures, the assessment of HOTS is becoming increasingly important. Moreover, AASs are in response to the emerging need of personalization while assessing HOTS.

Future work of the institution of the main author will address these issues by implementing a new AAS providing personalized assessment of not only LOTS, but also HOTS. The system designers will take advantage of the benefits of existing systems and compensate their disadvantages by taking into account more sophisticated feedback techniques and methods. This development will result in providing feedback

that is appropriate for the students' context, knowledge level, individual characteristics, preferences, behaviour and attentiveness. Thereby, the proposed feedback dimensions help identifying the potential of personalization that feedback actually has.

References

1. Anderson, L. W.; Krathwohl, D. R.; Airasian, P. W.; Cruikshank, K. A.; Mayer, R. E.; Pintrich, P. R.; et al. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Addison Wesley Longman, Inc.
2. Benyon, D.; Turner, P. & Turner, S. (2005). *Designing Interactive Systems*. People, Activities, Contexts, Technologies, Pearson Education, Ltd. Edinburgh.
3. Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), (pp. 7-74)
4. Bloom, B. S.; Engelhart, M. D.; Furst, E. J.; Hill, W. H. & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives*, Handbook 1: Cognitive Domain. Longman.
5. Brusilovsky, P.; Eklund, J. & Schwarz, E. (1998). Web-based Education for All: A Tool for Development Adaptive Courseware. *Computer Networks and ISDN Systems*, 30(1-7), (pp. 291-300). Elsevier.
6. Brusilovsky, P.; Sosnovsky, S. & Shcherbinina, O. (2004). *Quizguide: Increasing the Educational Value of Individualized Self-Assessment Quizzes with Adaptive Navigation Support*. Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 1806-1813.
7. Cheniti-Belcadhi, L.; Henze, N. & Braham, R. (2008). Assessment personalization in the Semantic Web. *Journal of Computational Methods in Sciences and Engineering*, 8(3), (pp. 163-182)
8. Chuang, S. & O'Neil, H. F. (2006). *Role of task-specific adapted feedback on a computer-based collaborative problem-solving task*. Web-based learning: Theory, research.
9. Conejo, R.; Guzmán, E.; Millán, E.; Trella, M.; Pérez-De-La-Cruz, J. L. & Rios, A. (2004). SIETTE: A Web-based Tool for Adaptive Testing. *International Journal of Artificial Intelligence in Education*, 14(1), (pp. 29-61)
10. Conlan, O. (2005). *The Multi-Model, Metadata Driven Approach to Personalised eLearning Services*.
11. Corbett, A. T. & Anderson, J. R. (2001). *Locus of feedback control in computer-based tutoring: impact on learning rate, achievement and attitudes*. Proceedings of ACM Conference on Human Factors in Computing Systems (pp. 245-252).
12. Czerwinski, M. P. & Larson, K. (2003). Cognition and the Web: moving from theory to Web design. *Human Factors and Web Development*, (pp. 147-165)
13. De Bra, P. & Calvi, L. (1998). AHA! An open adaptive hypermedia architecture. *New Review of Hypermedia and Multimedia*, 4(1), (pp. 115-139)
14. Dempsey, J. V. & Wager, S. U. (1988). A Taxonomy for the Timing of Feedback in Computer-Based Instruction. *Educational Technology*, 28, (pp. 20-25)
15. Fadel, C., Honey, M. and Pasnik, S. (2007). Assessment in the Age of Innovation. In *Education Week*, 26(38), (pp. 34-40)
16. Fisher, R. (2006). Thinking Skills. In A. J., T. Grainger, & D. Wray (Eds.), *Learning to Teach in the Primary School* (pp. 374-386). Routledge.
17. Gouli, E.; Papanikolaou, K. & Grigoriadou, M. (2002). *Personalizing Assessment in Adaptive Educational Hypermedia Systems*. Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 153-163).
18. Hambleton, R. K.; Swaminathan, H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory* (p. 184). Sage Publications, Inc.
19. Iahad, N. & Dafoulas, G. A. (2004). *The Role of Feedback in Interactive Learning Systems: A Comparative Analysis of Computer-Aided Assessment for Theoretical and Practical Courses*. Proceedings of the IEEE International Conference on Advanced Learning Technologies (pp. 535-539).
20. IEEE LTSC PAPI (2001). *Public and Private Information*, <http://www.cen-ltso.net/Main.aspx?put=230>
21. IMS LIP (2005). *Student Information Package*, <http://www.imsglobal.org/profiles>
22. IMS QTI (2006). *Question and Test Interoperability*, <http://www.imsglobal.org/question>
23. King, F. J.; Goodson, L. and Rohani, F. (1998). *Higher Order Thinking Skills*, Retrieved January 31, 2011, from http://www.cala.fsu.edu/files/higher_order_thinking_skills.pdf
24. Kozloff, M.A. and Wilmington N.C. (2002). *Three requirements of effective instruction: Providing sufficient scaffolding, helping students organize and activate knowledge, and sustaining high engaged time*.
25. Kulhavy, R.W. & Anderson, R.C. (1972). Delay-retention effect with multiple-choice tests In *Journal of Educational Psychology*, Volume 63 (pp. 505-512)
26. Kulhavy, R.W. & Stock, W. A. (1989). Feedback in Written Instruction: The Place of Response Certitude. *Educational Psychology Review*, 1(4), 279-308.
27. Kulhavy, R. W. & Wager, W. (1993). *Feedback in programmed instruction: Historical context and implications for practice*. Interactive instruction and feedback (pp. 3-20). Educational Technology Publications.
28. Lalos, P.; Retalis, S. & Psaromiligkos, Y. (2005). *Creating personalised quizzes both to the learner and to the access device characteristics: the Case of CosyQTI*. Proceedings of the Workshop on

- Authoring of Adaptive and Adaptable Educational Hypermedia (pp. 1-7).
29. Lazarinis, F.; Green, S. & Pearson, E. (2009). Focusing on content reusability and interoperability in a personalized hypermedia assessment tool. *Multimedia Tools and Applications*, 47(2), (pp. 257-278)
 30. Lewis, A. and Smith, D. (1993). Defining Higher Order Thinking. In *Theory into Practice*, 32(3), (pp. 131-137)
 31. Lütticke, R. (2004). *Problem Solving with Adaptive Feedback*. Proceedings of the Adaptive Hypermedia and Adaptive Web-based systems Conference (AH 2004) (pp. 417-420).
 32. Mason, B. J. & Bruning, R. (2001). *Providing feedback in computer-based instruction: What the research tells us*. Retrieved June 29, 2010, from <http://dwb.unl.edu/Edit/MB/MasonBruning.html>
 33. Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (pp. 745-783). Mahwah: Lawrence Erlbaum Associates.
 34. Osborne, C. & Winkley, J. (2006). *Developments in On-Screen Assessment Design for Examinations*. Proceedings of the 10th CAA International Computer Assisted Assessment Conference (pp. 343-358).
 35. Pitkow, J. E. & Recker, M. M. (1995). Using the Web as a Survey Tool: Results from the Second WWW User Survey. *Computer Networks and ISDN Systems*, 27(6), (pp. 809-822)
 36. Sharp, H.; Rogers, Y. & Preece, J. (2007). *Interaction Design. Beyond Human-Computer Interaction*, 2nd Edition. John Wiley & Sons, Ltd. Chichester.
 37. Shepherd, E. and Godwin, J. (2004). *Assessments through the Learning Process*, Retrieved December 3, 2010, from http://www.questionmark.com/catalog/us/resources/Assessments_Through_the_Learning_Process.pdf
 38. Sugrue, B. (1995). A Theory-Based Framework for Assessing Domain-Specific Problem-Solving Ability. In *Educational Measurement: Issues and Practice*, 14(3), (pp. 29-35)
 39. Der Linden, W. J. van; & Glas, C. A. W. (2000). *Computerized Adaptive Testing: Theory and Practice* (p. 336). Springer Netherlands.
 40. Vasilyeva, E.; Puuronen, S.; Pechenizkiy, M. & Rasanen, P. (2007). Feedback Adaptation in web-based Learning Systems. *International Journal of Continuing Engineering Education and Life-Long Learning*, 17(4/5), (pp. 337-357)
 41. Wainer, H.; Dorans, N. J.; Eignor, D.; Flaughner, R.; Green, B. F.; Mislevy, R. J.; et al. (2000). *Computerized Adaptive Testing: A Primer* (p. 335). Lawrence Erlbaum Associates.
 42. Wuttke, H.-D.; Ubar, R.; Henke, K. and Jutman, A. (2008). *The synthesis level in Blooms Taxonomy a nightmare for an LMS*. In Proceedings of the 19th EAEEIE Annual Conference, 199-204.
 43. XML Topic Maps (2001). *XML Topic Maps Specification*, <http://topicmaps.org/xtm/>