PERSONALIZED ASSESSMENT OF HIGHER-ORDER THINKING SKILLS

Christian Saul
Fraunhofer Institute Digital Media Technology, Business Area Data Representation & Interfaces, Ilmenau, Germany

Heinz-Dietrich Wuttke
Faculty Informatics and Automation, Ilmenau University of Technology, Ilmenau, Germany

Keywords: Assessment, Adaptive Assessment Systems, Higher-order Thinking Skills, Simulations.

Abstract: As our society moves from an information-based to an innovation-based environment, it is not just important what you know, but how you can use your own knowledge in order to solve problems and create new knowledge. Hence, assessment systems need to evaluate not just the students' factual knowledge, but also their problem-solving and reasoning strategies. This leads to a demand for the assessment of higher-order thinking skills (HOTS). This paper analyzes HOTS and possibilities for their measurement. As a result, adaptive assessment systems (AASs) are in response to the emerging need of personalization while assessing HOTS. AASs take student's individual context, prior knowledge and preferences into account in order to personalize the assessment. This personalized support helps students develop HOTS. But, this paper also reveals several arising issues, which need to be addressed when measuring HOTS with AASs.

1 INTRODUCTION

Today, learning occurs in a variety of places not only within a teacher-student relationship, but also at home, work and through daily interactions with today's society. Whatever the environment of learning and method of delivery, it is crucial to obtain evidence about the knowledge, skills, attitudes they have in fact learned. Hence, the effective use of assessment is an integral part of developing successful learning materials and a critical catalyst for student learning (Conole & Warburton, 2005). Assessment is defined as a 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. Alongside several advantages, however, 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 adaptive assessment system (AAS) poses one way to realize personalization in online-assessments. It takes the student's individual context, prior knowledge and preferences into account in order to personalize the assessment, which may result in more objective assessment findings. But, as our society moves from an information-based to an innovation-based environment (Fadel et al., 2007), it is not just important what you know, but how you can use it in order to solve problems and create new knowledge. Hence, assessment systems need to evaluate not just the students' factual knowledge, but also their problem-solving and reasoning strategies. This leads to a demand for the assessment of higher-order thinking skills (HOTS). The focus of this paper is to analyze HOTS and to identify possibilities for their measurement. Hence, four established AASs (SIETTE, PASS, CosyQTI and iAdaptTest) will be examined in this respect.

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

Saul C. and Wuttke H..
PERSONALIZED ASSESSMENT OF HIGHER-ORDER THINKING SKILLS.
DOI: 10.5220/0003480204250430
In *Proceedings of the 3rd International Conference on Computer Supported Education* (ATTeL-2011), pages 425-430
ISBN: 978-989-8425-50-8

and knowledge consumer.

The remainder of the paper is organized as follows: The second chapter analyzes HOTS and tries to find an appropriate categorization. The third chapter looks at the assessment of these skills and associates AASs in this respect. For that reason, chapter four deals with AASs and their possibilities to assess HOTS. The findings are discussed in chapter five. Concluding remarks, future work and references complete the paper.

2 HIGHER-ORDER THINKING SKILLS

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). Hence, assessment systems need to evaluate not just the students' factual knowledge, but also their problem-solving and reasoning strategies, which are currently left to oral examinations or project work. These advanced thinking skills are known under the term HOTS.

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. HOTS are grounded in lower-order thinking skills (LOTS) such as simple application and analysis and linked to prior knowledge (King et al., 1998).

Thinking skills were conceptualized in a number of ways and at present there is little consensus with regard to the actual term. For a comprehensive overview, reference is made to King et al. (1998).

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 dissect 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. 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 1).

Table 1: Taxonomies of the Cognitive Domain.

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

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. Factual knowledge is knowledge that is essential to specific disciplines. 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 and meta-cognitive knowledge is knowledge of cognition in general as well as awareness of one's own cognition.

Reasons for selecting Anderson and Krathwohl's revision of Bloom's taxonomy as preferred basis for further understanding are the use of recent advancements in psychological and educational research (for example, constructivism, meta-cognition and self-regulated learning) and their general applicability in all subject matters for specifying teaching objectives, activities and assessments.

3 ASSESSMENT OF HIGHER-ORDER THINKING SKILLS

Assessment is regarded by many 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: (1) selection, (2) generation and (3) explanation. Selection means using simple question types such as multiple-choice and matching for identifying the most plausible assumption or the most reasonable inference. But, although multiple-choice questions can be used for separately measuring some specific HOTS such as deduction, inference and prediction (Bloom et al., 1956), they are inappropriate for measuring skills on the evaluation and creation level. Generation means using advanced question types, which let students more creativity in answering, such as free-text answers, essays and interactive and simulative tools (ISTs) for measuring HOTS. Mitchell et al. (2002) proposed a software system (AutoMark) for evaluating free-text answers to open-ended questions. AutoMark uses the techniques of information extraction to provide computerized marking of short free-text responses. The technique of automatically evaluating essays is used by Burstein and Marcu (2003) in a system called E-Rater. E-Rater identifies thesis and conclusion statements from student essays on six different topics. Furthermore, ISTs can deal with complex real-life problems that require students to employ a number of HOTS in order to solve them. This coincides with Bennett (1998) as his vision of assessment. He pointed out that assessment has not yet achieved its full potential and predicted a dramatically improvement in using simulation and virtual reality while assessment. The vision of Bennett was followed up by Cleave-Hogg et al. (2000) and Wuttke et al. (2008) in specialized trainings and assessment tools. Cleave-Hogg et al. used an anesthesia simulator for assessing medical students' performance while narcosis. Students were given patient information and expected to apply their knowledge, demonstrate the necessary technical skills and use professional judgment. Wuttke et al. proposed two concepts for learning-by-doing: a remote laboratory where students can design, verify and implement digital circuits and control systems and a collection of interactive tools. Using these tools, the students can explore their knowledge and get new ideas. As computer video games are highly virtual interactive environments as well, they have become interesting to educators and researchers over the last decade. Rice (2007) analyzed different video games and their potential in addressing HOTS and provided a tool, which will assist educators in deciding what video games to use with their students. Moreover, portfolios were also recommended for measuring HOTS (Lankes, 1995). Finally, 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 order to ensure the validity of the responses, Norris (1989) recommended a thinking-aloud procedure. This enables identifying when correct responses were chosen through faulty thinking or incorrect responses through valid thinking.

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 addressing 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. Furthermore, valid assessment of HOTS requires that students are unfamiliar with the questions or tasks they are asked to answer or perform.

In this regard, a demand towards personalization arises to take care of the individual of the students. In the context of information and communication technologies, personalization can be defined as the process of tailoring something to individual characteristics, preferences and abilities. One way to realize personalization in assessments are adaptive assessment systems.

4 ADAPTIVE ASSESSMENT SYSTEMS

Several AASs and technologies exist, which can be used to test students at their current knowledge level and change their behavior and structure depending on the students' previous responses, individual context, prior knowledge and preferences. There are two types of techniques that can be applied in AASs namely adaptive testing (Wainer et al., 2000; Van der Linden & Glas, 2000) and adaptive questions (Pitkow & Recker, 1995).

4.1 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 to estimate the probability of a correct answer to a particular question and to select an appropriate question accordingly. 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, but has some disadvantage in terms of estimating students' knowledge level separated to the particular topics in a test.

PASS (Personalized ASSessment) is a web-based assessment module, which estimates students' performance through multiple assessment options tailored to students' responses. Advantageous of PASS is the consideration of the students' navigational behavior, 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.

4.2 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. 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, IMS LIP and IEEE LTSC PAPI learning standards, which makes the system interoperable with other standard-compliant learning tools and systems. Regarding the authoring of questions, the limited rule system and the few question types restrict the incorporation of didactic philosophy and methods.

iAdaptTest is a desktop-based modularized adaptive testing tool conforming to the IMS QTI, the IMS LIP and XML Topic Maps in order to improve the reusability and interoperability of the data. But, iAdaptTest provides only a few question types and the implemented feedback and help is rather simple and does not enable personalized support.

4.3 Comparison Towards the Assessment of Higher-order Thinking Skills

This chapter will analyze and compare the previous described AASs with respect to the assessment of HOTS. According to chapter 3, 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, special attention was laid on these criteria. The results of the comparison are provided in Table 2.

Table 2: Comparison of SIETTE, PASS, CosyQTI and iAdaptTest towards the Assessment of HOTS.

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

The table above shows that all AASs are 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.

Summarized this means that all analyzed AASs can only be used in measuring at least some HOTS. The potential of these AASs for assessing thinking skills is presented in Table 3. 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 3: Taxonomy Matrix of SIETTE, PASS, CosyQTI and iAdaptTest.

		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)		

As mentioned in chapter 3, the assessment of HOTS should not only adapt for diverse student needs, but also should support the students in retrieving their prior knowledge necessary for using HOTS. This facet of providing personalized feedback in AASs was already investigated in earlier research (Saul et al., 2010). The results have shown that each of the AASs provides possibilities to incorporate feedback in the assessment process, but the use of feedback techniques is limited.

5 DISCUSSION

In the last chapters, the importance of HOTS and their assessment was emphasized. AASs are in response to the emerging need of personalization while assessing HOTS. But as shown, the incorporation of advanced question types is very poor, even though it is a prerequisite in assessing HOTS (see chapter 3). Another demand towards the assessment of HOTS is personalized support of the students in developing their HOTS. As mentioned, personalization of feedback is still insufficiently implemented or even not addressed in these systems. What is missing is an AAS that incorporates ISTs in order to bridge the gap between the assessment of HOTS and adaptive assessment. It is not just a case of allowing the IST to exist within the system, but of allowing the AAS and the IST to communicate at a much deeper level to enable more efficient and effective personalized assessments of students' HOTS.

This raises issues about the communication between both systems. It needs to be specified what is communicated, when and how. Further issues concern where the accuracy of the student activity should be assessed, where the questions should be marked, where the feedback come from, where the results should be reported, whether the state should be preserved, etc. (Thomas et al., 2004). These questions need to be aligned with the application scenarios taken into account and strongly influence the design of the communications interfaces.

More substantial is the level of integration of AAS and IST. Thomas et al. (2005) proposed three levels of integration between assessment system and IST ranging from no communication up to two-way communication to set up and mark questions.

Another issue concerns the technique used for building the AAS. The technique of adaptive testing is restricted to multiple-choice questions. In contrast, the technique of adaptive questions is more flexible in this respect and not restricted to any question types.

6 CONCLUSIONS AND FUTURE WORK

The objective of this paper was to analyze HOTS and to identify possibilities for their measurement. The analysis was caused by the understanding of evaluating not just the students' factual knowledge, but also their problem-solving and reasoning strategies, which is currently left to oral examinations or project work. In today's society, it is not just important what you know, but how you can use it in order to solve problems and create new knowledge. The results of the analysis pointed out those AASs are in response to the emerging need of personalization while assessing HOTS. They take student's individual context, prior knowledge and preferences into account in order to personalize the assessment. But, the other way around, the assessment of HOTS in the analyzed AASs (SIETTE, PASS, CosyQTI and iAdaptTest) is still insufficiently implemented or even not addressed. As an example, the incorporation of advanced question types is very poor, even though it is a prerequisite in assessing HOTS. In addition, the paper also revealed several arising issues, which need to be addressed when measuring HOTS with AASs.

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. This will be realized by incorporating ISTs in a holistic assessment process.

REFERENCES

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., et al., 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, Addison Wesley Longman, Inc.

Bennett, R. E., 1998. Reinventing Assessment. Speculations on the Future of Large-Scale Educational Testing, Policy Information Center.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H. and Krathwohl, D. R., 1956. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*, Longman.

Burstein, J., and Marcu, D., 2003. A Machine Learning Approach for Identification Thesis and Conclusion Statements in Student Essays. In *Computers and the Humanities, 37(4)*, 455-467.

Cleave-Hogg, D., Morgan, P. and Guest, C., 2000. Evaluation of Medical Students' Performance in Anaesthesia Using a CAE Med-Link Simulator System. In *Proceedings of the Fourth International Computer Assisted Assessment Conference*, 119-126.

Conejo , R., Guzman, E., Millan, E., Trella, M., Perez-de-la-Cruz, J. and Rios, A., 2004. SIETTE: A Web-Based Tool for Adaptive Testing, In *International Journal of Artificial Intelligence in Education, 14*, 29-61.

Conole, G. and Warburton, B., 2005. A review of computer-assisted assessment, In *Alt-J, Research in Learning Technology, 13(1)*, 17-31.

Fadel, C., Honey, M. and Pasnik, S., 2007. Assessment in the Age of Innovation. In *Education Week, 26(38),* 34-40.

Gouli, E., Papanikolaou, K. and Grigoriadou, M., 2002. Personalizing assessment in adaptive educational hypermedia systems. In *Proceedings of the Second International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, 153-163.

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

IEEE PAPI, 2002. Public and Private Information, http://www.cen-ltso.net/Main.aspx?put=230

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

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.

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.

Lalos, P., Retalis, S. and Psaromiligkos, Y., 2005. Creating personalised quizzes both to the learner and to the access device characteristics: the Case of CosyQTI. In *Proceedings of Workshop on Authoring of Adaptive and Adaptable Educational Hypermedia*, 1-7.

Lankes, A. M. D., 1995. Electronic portfolios: A new idea in assessment, Retrieved February 10, 2011, from www.eric.ed.gov.

Lazarinis, F., Green, S. and Pearson, E., 2009. Focusing on content reusability and interoperability in a personalized hypermedia assessment tool. In *Multimedia Tools and Applications, 47(2)*, 257-278.

Lewis, A. and Smith, D., 1993. Defining Higher Order Thinking. In *Theory into Practice, 32(3)*, 131-137.

Mitchell, T., Russell, T., Broomhead, P. and Aldridge, N., 2002. Towards robust computerised marking of free-text responses. In *Proceedings of the 6th CAA Conference*, 233-249.

Norris, S., 1989. Can we test validly for critical thinking? In *Educational Researcher, 18(9)*, 21-26.

Pitkow, J. and Recker, M., 1995. Using the Web as a survey tool: results from the second WWW user survey, In *Computer Networks and ISDN Systems*, *27*, 809-822.

Rice, J. W., 2007. Assessing higher order thinking in video games. In *Journal of Technology and Teacher Education, 15(1)*, 87-100.

Sangwin, C. J., 2003. Assessing higher mathematical skills using computer algebra marking through AIM. In *Proceedings of the Engineering Mathematics and Applications Conference (EMAC03)*, 229-234.

Saul, C., Runardotter, M. and Wuttke, H.-D., 2010. Towards Feedback Personalisation in Adaptive Assessment, In *Proceedings of the Sixth EDEN Research Workshop*.

Shepherd, E. and Godwin, J., 2004. Assessments through the Learning Process, Retrieved December 3, 2010, from http://www.questionmark.com/us/whitepapers.

Sugrue, B., 1995. A Theory-Based Framework for Assessing Domain-Specific Problem-Solving Ability. In *Educational Measurement: Issues and Practice, 14(3)*, 29-35.

Thomas, R., Ashton, H., Austin, B., Beevers, C., Edwards, D. and Milligan, C., 2004. Assessing Higher Order Skills Using Simulations. In *Proceedings of the 8th CAA Conference*.

Thomas, R., Ashton, H.S., Austin, B., et al., 2005. Cost effective use of simulations in Online Assessment. In *Proceedings of the 9th CAA Conference*.

Van der Linden, W. and Glas, C., 2000. *Computerized adaptive testing: Theory and practice*, Springer Netherlands.

Wainer, H., Dorans, N., Eignor, D., Flaugher, R., Green, B. and Mislevy, R., 2000. *Computerized Adaptive Testing: A Primer*, Lawrence Erlbaum Associates.

Wuttke, H.-D., Ubar, R., Henke, K., Jutman, A., 2008. The synthesis level in Blooms Taxonomy a nightmare for an LMS. In *Proceedings of the 19th EAEEIE Annual Conference*, 199-204.