Learning Outcomes and a Taxonomy as a Starting Point for creating digital Multiple-choice Tests

Bjørn Klefstad

Assistant Professor
Department of Informatics and e-learning
Sør-Trøndelag University College
bjorn.klefstad@hist.no

Geir Maribu

Associate Professor
Department of Informatics and e-learning
Sør-Trøndelag University College
geir.maribu@hist.no

Svend Andreas Horgen

Assistant Professor
Department of Informatics and e-learning
Sør-Trøndelag University College
svend.horgen@hist.no

Thorleif Hjeltnes

Associate Professor
Department of Informatics and e-learning
Sør-Trøndelag University College
thorleif.hjeltnes@hist.no

Abstract

The use of digital multiple-choice tests in formative and summative assessment has many advantages. Such tests are effective, objective, and flexible. However, it is still challenging to create tests that are valid and reliable. Bloom's taxonomy is used as a framework for assessment in higher education and therefore has a great deal of influence on how the learning outcomes are formulated. Using digital tools to create tests has been common for some time, yet the tests are still mostly answered on paper. Our hypothesis has two parts: first, it is possible to create summative tests that match different levels and learning outcomes within a chosen subject; second, a test tool of some kind is necessary to enable teachers and examiners to take a more proactive attitude to(wards) different levels and learning outcomes in a subject and so ensure the quality of digital test designing. Based on an analysis of several digital tests we examine to what degree learning outcomes and levels are reflected in the different test

questions. We also suggest functionality for a future test tool to support an improved design process.

Keywords: Bloom's taxonomy, digital test, exam, learning outcome, summative and formative assessment, test tool.

Introduction to digital tests

A digital test is a collection of questions that are designed, answered and graded using digital tools in order to measure competence within a given subject. Paper-based tests have (conventionally) been used for assessment, especially when it comes to final exams and courses with a large number of students, because they fulfill the requirements for objective and effective grading (Bjørgen and Ask, 2006). Digital tests offer extended possibilities for testers and test-takers, including several types of questions that differ from the traditional multiple-choice type. Examples of these new types of questions are "drag and drop," "order a sequence" and "fill-in-the-blank." In addition, it is possible to use pictures, video, audio and links in the questions and in the answer alternatives. This makes it possible to ask more advanced questions and test competencies on various levels.

The Quality Reform for higher education in Norway lists several requirements for assessment methods with a special focus on continuous assessment (Kvalitetsreformen 2001). Learning Management Systems (LMS) like Fronter, it's learning and Moodle offer basic digital test tools. The introduction of LMS tools in the area of higher education has turned the focus towards digital tests as a usable assessment method not only for summative assessments (exam) but also for formative assessment (i.e. continuous assessment). Increasing numbers of teachers use digital tests in formative assessment. The benefits of digital tests are automatic grading, instant feedback, motivation, repetition and variation. Many teachers regard the increased efficiency and students' familiarity with such tests during formative use as motivation for also using digital tests in summative assessments.

The challenges of designing digital tests are many. It is not easy to design high-quality questions, and multiple-choice tests are often criticized for testing basic knowledge only. Sirnes (2005) argues for the importance of designing tests that are both valid and reliable. Validity is a measure for the degree to which the test measures what it is meant to measure, regarding both subjects and cognitive level. Reliability is a measure for how reliable the test is, i.e. to what degree we can trust the test and if it gives the same results when used again. There is a close connection between the form of the assessment, the learning process and the motivation. Brooks (2002) emphasizes the effect of WYTIWYG (What You Test Is What You Get) that says assessment and learning are closely connected. When you have chosen an assessment method you have indirectly chosen a kind of learning.

To use digital tests successfully for formative and summative assessments, we must strive for high quality in the design of the tests. The learning outcomes define the subjects and at what level the subjects should be learnt. Our hypothesis: To ensure the quality of the test it is necessary to focus on the level and learning outcomes for every question. Current test tools do not have a functionality that clearly shows this connection between the questions and the learning outcomes and taxonomy level. In this paper we present a method to help the teacher design precise tests and also suggest how a digital test tool can be used to reach this goal.

Learning outcomes

A learning outcome is a description of what a student should know after fulfilling a given course. That is what the student should *know, understand* and *be able to demonstrate* on completing the course. Until now courses have been described through a list of content and objectives and nothing more. Such a description is mainly for the benefit of teachers and shows the institution's intentions for the course.

Learning outcomes tell us what the student should be able to do after fulfilling the course, not the intentions of the institution and the teacher. Learning outcome is a description of the knowledge, comprehension and skills the student should have obtained during the course. Learning outcomes are a student-centered way of describing the course. Learning outcomes are about learning, while objectives are about teaching. In this way, learning outcomes mirror the switch from teacher-centered teaching to student-centered learning.

The concept of 'learning outcomes' is well documented in the literature about the Bologna process (Bologna UK-website 2009). That means the European educational ministers see learning outcome descriptions as a useful tool to implement a more transparent European educational system across international borders. Thorough descriptions of learning outcomes make it much easier for institutions to approve courses from other educational institutions in European countries, and thus increase student mobility.

Learning outcomes have a specific structure, consisting of a verb and an object. The verb tells us what the student should be capable of doing after the course and the object tell us in what subject theme he/she can do it. A typical learning outcome is written in the following form: "after fulfilling this course you should manage to: separate gill mushrooms from pore mushrooms," "explain how mushrooms breed," "list the names of the six most consumable mushrooms" and "distinguish between the six most consumable mushrooms."

In the above examples we often see verbs like "distinguish," "explain," "list" and "identify." There are many verbs that can be used in learning outcome descriptions. The verbs tell us what kind of knowledge will be attained and at what level, for example, basic knowledge or comprehension. Typical verbs used for basic knowledge are *define, list, describe, recognize, name* etc. For the comprehensive level, verbs like *explain, match, summarize, distinguish* and *locate* are typical.

Taxonomies

A learning outcome is not a new concept. As early as 1956, Benjamin Bloom at University of Chicago published a list of learning outcomes (or "learning goals," as he called them). This list is well known in the educational system and has experienced a renaissance in recent years, as a result of greater interest in learning outcomes among higher education institutions.

The list is known as "Bloom's taxonomy" and classifies the learning outcomes that the teacher can create for students. Bloom realized that there are three domains of learning: *cognitive*, *affective* and *psychomotor*. Each domain is divided into several levels from the basic (level) to the complex. In this paper we look only at the cognitive domain.

Cognitive levels deal with knowledge and the development of intellectual skills, restating and recognition of facts, procedures and concepts. The six levels in the cognitive domain are:

- Knowledge
- Comprehension
- Application
- Analysis
- Synthesis
- Evaluation

In the mid-1990s, Anderson (2001) made some changes to Bloom's cognitive learning taxonomy. The most important change was renaming the six levels, from nouns to verbs, resulting in a more active formulation of learning goals (or learning outcomes, as we say today).

- *Knowledge* was replaced with to *reproduce*
- *Comprehension* was replaced with to *comprehend*
- *Application* was replaced with to *apply*
- *Analysis* was replaced with to *analyze*
- *Synthesis* was replace with to *create*
- *Evaluation* was replaced with to *evaluate*

This reformulated taxonomy is more in line with the new thinking about learning outcomes. Now we discuss actions, telling what has to be done to learn something, and not only describing what we wish to obtain.

Learning outcome descriptions can fit readily into Bloom's levels and are supplied with verbs to describe the learning goals for each level. Typical verbs for the levels in Bloom's taxonomy in the cognitive domain are:

- *Knowledge*: define, describe, identify, list, label, recite, name.
- *Comprehension*: explain, group, distinguish, match, restate, extend, summarize, give examples.
- *Application*: prepare, produce, choose, apply, solve, show, sketch, relate, use, generalize.
- *Analysis*: analyze, compare, differentiate, subdivide, classify, categorize, survey, point out.
- *Synthesis*: compose, develop, design, combine, construct, create, plan, invent, organize.
- *Evaluation*: judge, evaluate, criticize, consider, recommend, support, relate, decide, discuss.

Using these active verbs, we can abandon passive phrases like "have more knowledge in," "give an understanding of" etc.

In the book *Teaching for Quality Learning at University* (Biggs 2007), Biggs uses learning goals and levels to improve learning. Biggs uses the name "outcome-based teaching and learning," in which the focus falls on how to design good, descriptive and active goals for learning in courses. Biggs uses the name "learning outcome" instead of "learning goal" because he wants to describe what the student should have achieved after the given learning activity is finished. With a bundle of learning outcome descriptions for each learning module, it is easy to measure what the student has learnt by reformulating the learning outcomes to questions in an assessment (for example, in an exam). And even better, the assessment questions can be made simultaneously with the learning outcomes, i.e. before the course or learning module starts.

Methodology

Teachers with some experience of using digital tests received a one-hour introduction to Bloom's taxonomy and then received instructions for how to analyze their own digital tests in the context of Bloom's cognitive levels. The

procedure was to make a subjective evaluation of the tests they have designed, and classify the questions according to which Bloom level they belonged. The result of this analysis is a table showing the number of questions in each of the six Bloom levels.

The teachers were free to choose which tests they would include. The only requirement was that the test should be of the multiple-choice variety.

The tests were used at the Bachelor in Information Technology at Sør-Trøndelag University College in these courses: "Information Security Management," "Programming in Java," "Informatics 2" (with these parts: Operating Systems, LAN Management and Data Communication), "The ITIL standard," "Practical Linux" and "Ajax programming."

Results and analysis

The results in the table below show a significant predominance of questions from the knowledge level (over 50%) for the chosen tests. Some questions are classified as comprehension, i.e. Bloom's second level. In the third level (application) only a few questions are found.

Bloom levels Course	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Total number of questions
Information security management	25						25
Programming in Java	16	8	1				25
Informatics 2	52	2					54
The ITIL standard	40						40
Practical Linux	51	38	3				92
Ajax programming	21	7	6	4	1	1	40

Figure 1: Number of questions for each Bloom level in six analyzed tests. An empty cell means no questions at this level.

The teachers had varied experience of designing multiple-choice tests, and therefore their subjective evaluations as to which level a question belonged to is very likely also variable. There is also some variation across courses in how necessary and challenging it is to design questions for the upper levels of Bloom's taxonomy. In summative assessments, multiple-choice tests can be used in combination with other assessment methods. Therefore, teachers may prefer to design multiple-choice tests with all questions at the knowledge level because they intend to use other assessment methods to test the remaining Bloom levels.

In the course "Ajax programming," we see that all Bloom levels are used. In addition, all questions in this test are given variable weighting, in contrast to the other five tests where each question is given one point. That means if we count the number of points instead of the number of questions for each level, the first five tests in the table will be unchanged, while the test for "Ajax programming" includes 23 points at the knowledge level, 9 points at the comprehension level, 9 points at the application level, 10 points at the analysis level, 2 points at the synthesis level and 1 point at the evaluation level, giving 54 points in total. Therefore this test differs significantly from the other tests, since the teacher is aware that questions at higher levels are more difficult to answer and a correct answer should therefore be more highly rewarded.

The teachers behind these tests have used multiple-choice tests for several years. The teacher making the test for the course "Ajax programming" shows a high awareness of Bloom's taxonomy in designing the test, because he distributed the questions among all the Bloom levels. We have not estimated how valid and how reliable the chosen tests are, but pending a more thorough analysis it appears as though awareness of Bloom levels will contribute to increased test quality.

Our analysis indicates that the questions in digital tests do not usually cover every level in Bloom's taxonomy. However, if the teacher is aware of the taxonomy levels they may find it easier to include them, and we think that some procedures and tools for the process of designing tests will further aid teachers in creating good tests.

A procedure for designing good and valid digital multiple-choice tests

We usually want to distinguish between competencies by using questions with different degrees of difficulty. In higher educational institutions Bloom's taxonomy has been used to classify assessments according to the different cognitive levels: knowledge, comprehension, application, analysis, synthesis and evaluation. With reference to Bloom's taxonomy, it is more demanding to comprehend something than to reproduce some knowledge (Imsen 2002:208), and it is important to assess students at various levels.

Biggs argued for the importance of presenting the learning outcomes clearly to students, as well as of offering learning activities that build on the learning outcomes, and finally, of assessing students using the learning outcomes (Biggs 2007). The learning outcomes should therefore be given a key role during the design of the test.

We see the need for an increased focus on providing questions at all Bloom levels and covering all learning outcomes in tests. Our suggestion for a procedure to make tests better is first to design detailed, accurate learning outcomes for each course, and next to connect the tests to a chosen taxonomy, Bloom's for example. This connection will make it much easier to achieve the correct perspective, create an awareness of what we are testing and to what level, and in this way increase the quality of the tests we design.

Our hypothesis is that a good and valid test contains questions that cover various learning outcomes and taxonomy levels. If the teacher designing the test can specify which learning outcome each question fits into, and into which cognitive level, the quality of the test as a whole will increase. Such a classification is subjective, so will have the greatest degree of consistency if the learning outcomes are well defined and if help (from a tool) is offered to classify the level of the questions. If each question is classified as we suggest, it is possible to create a number of questions for each learning outcome and level in a clearly arranged table.

The autumn 2009 final exam in the course "Web-programming using PHP" at Sør-Trøndelag University College was arranged on the Internet as a digital test, weighted at 1/3 of the course; program development made up the other 2/3 of the students' grades. The course has seven primary learning outcomes and about 60 secondary and more detailed learning outcomes. The table below shows how the test questions are distributed among the various learning outcomes and the levels in Bloom's taxonomy. Since this was a final exam, the primary learning outcomes are used in the classification.

Learning outcome / Bloom levels	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	Sum
LO1: Solve problems using PHP-scripts as a tool			2 (3)				2 (3)
LO2: Use PHP to create large web-based systems		1 (1)					1 (1)
LO3: Structure the program code in a good way	1 (1)		2 (3)				3 (4)
LO4: Find and implement working security actions for a web-solution	2 (2)						2 (2)
LO5: Create usable and functional systems							
LO6: Program towards a remote database							
LO7: Explain how a state can be saved with PHP	1 (1)	1 (1)					2 (2)
None	1 (1)	3 (4)		2 (3)			6 (8)
Sum	5 (5)	5 (6)	4 (6)	2 (3)			16 (20)

Figure 2: Table showing the distribution of questions (number of points in parenthesis) for each learning outcome and for various levels in Bloom's taxonomy.

In the above table we can see that learning outcome LO1 has two questions on the level of "application." Since one of these two questions counts as two points, there are three points in total for LO1/application. Learning outcome LO3 has one question on the level "knowledge" and two questions on the level "application," and one of the questions counts as two points. No questions cover learning outcome LO5 and LO6, and there is no question covering the levels "synthesis" and "evaluation."

A table like this (one) gives the teacher a quick and effective overview of what learning outcomes and taxonomy levels are being used in the test. Our hypothesis says that working in this manner will contribute to increase awareness of the quality of the test in regard to validity. The teacher can compensate for any shortcomings by creating more questions, or by seeing which learning outcomes must if necessary be covered by other assessment methods. The idea of the table is to show the big picture and to make the teacher aware of what the test will cover.

In the table we also find some questions in the category "none." These are usually questions that were created without thinking about learning outcomes, or in which the learning outcomes are too diffuse to be classified. Biggs (2007) emphasizes that learning outcomes should be well formulated and specific. If the teacher finds out that the question in the "none" category is still relevant for the test, the table reveals some shortcomings in the learning outcomes for the test (and for the course). We therefore have to create a new learning outcome, for example, LO8: "Master the syntax of the PHP language." In that case it would be wise to switch several of the questions from category "none" into category LO8. The test and the table is therefore a great help for the teacher in focusing on the important learning outcomes.

To make high quality tests that cover various learning outcomes and cognitive levels, it is crucial that the teacher receive the support and functionality of a test tool to classify the questions while working on creating a test. The important quality assurance is then completed during the design process of the test and not after a disappointing result.

Tools

In LMS systems (it's learning, Fronter, Moodle) currently used in Norway, good tools exist for designing digital tests. Digital test tools help create many question types, for example "drag and drop," "order in sequence," "fill in blanks" and more. In addition, several kinds of stimuli can be included in questions and alternatives, for example, text, images, video, audio, links and other digital content. Armed with all these possibilities, we can design advanced questions and alternatives and test various levels of competence in accordance with Bloom's taxonomy.

Based on our work, we suggest that a test tool should offer support for the design of questions, i.e. to conform with learning outcomes and Bloom levels, and not only focus on question and stimuli types. Existing test tools do not offer that kind of support.

At the Department of Informatics and e-learning at Sør-Trøndelag University College, we developed a tool called DigiTEST. This is a tool for designing, running and managing tests. We have implemented functionality to help the teacher control the learning outcomes and their connections to the chosen taxonomy. When the teacher designs a test, he begins by entering the learning outcomes into the tool. This procedure forces the teacher to decide which learning outcomes should be tested. The teacher next chooses a preferred taxonomy (for example, Bloom's). The test tool now offers an empty table that links learning outcomes and levels in the taxonomy. After defining the learning outcomes and chosen taxonomy, the teacher formulates the questions. In DigiTEST the teacher can choose a learning outcome and a category within the chosen taxonomy for each question.
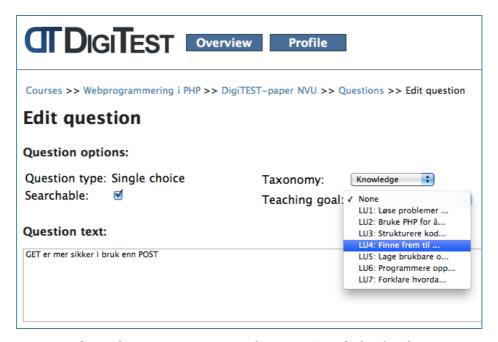

Figure 3: The teacher creates questions in the DigiTEST tool after first having registered learning outcomes and a chosen taxonomy level.

While the teacher is adding new questions, he can see a table showing the current state of the learning outcomes and taxonomy levels in the test, and the links between them. Here the teacher gets the complete picture of the test and can easily see which learning outcomes and levels are covered and which questions must be created in order to make the test as complete as possible. Figure 4 below shows such an example.

	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
LU1: Løse problemer...						
LU2: Bruke PHP for...						
LU3: Strukturere...	1		2			
LU4: Finne frem til...	1					
LU5: Lage brukbare...						
LU6: Programmere...						
LU7: Forklare...						

Figure 4: The teacher has so far created four questions. With DigiTEST, one can visualize learning outcomes and taxonomy levels.

We observed that without support from a tool, most tests only offer questions at the lower taxonomy levels, and not all learning outcomes are covered. It is our opinion that there is a strong need for a tool that helps teachers increase the quality of digital tests.

Conclusion and work still to be done

It is demanding to create a valid digital test. Our analysis of various tests shows that most of the questions end up at lower taxonomy levels. It is a goal for many educators to also test at higher levels, and test the complete content of the course. To ensure the quality and (the) validity of tests, we have seen a need for increasing teachers' awareness of the importance of classifying the questions in a taxonomy and for using learning outcomes.

One procedure to increase teachers' awareness of these aspects is to build this functionality into a test tool. Our way to emphasize this link is to let the teacher register the taxonomy level and the learning outcomes that each question covers. Next, the test tool can add up all the questions covering each of the learning outcomes for the various levels. Shown in a table, these numbers give the teacher an excellent overview of what learning outcomes are covered by the test and what learning outcomes are still uncovered.

We will continue to work with digital tests for the years to come. After we gain more experience with this tool and its awareness-raising method, we will more closely examine the extent to which the method contributes to increased competency in creating digital tests for use in educational institutions.

References

Analysis/data of (the) digital test used in this paper is available at http://digitest.no/diverse/paper-nvu-2010

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J. & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. Longman: New York (2001).

Biggs, J., & Tang, C. (2007). Teaching for quality learning at university. Buckingham: Open University Press/Society for Research into Higher Education (Third edition), (pp. 70-72), New York, McGraw-Hill.

Bjørgen, I.A. & Ask, H.. (2006). "Læringsmøter IV – Kvalitetsreformen ved Psykologisk institutt." *Rapport NTNU*, Psykologisk institutt.

Bloom B. (1956). "Taxonomy of Educational Objectives: Handbook 1, The cognitive domain". New York: McKay

Bologna UK-website (2009): http://www.dcsf.gov.uk/bologna

Brooks, V. (2002). "Assessment in Secondary Schools. The New Teacher´s guide to Monitoring, Assessment, Recording, Reporting and Accountability." Buckingham: Open University.

Dave, R.H. (1975). *Developing and Writing Behavioral Objectives.* In: Armstrong, R. J. (ed.). Tucson, Arizona: Educational Innovators Press.

Fjørtoft, H. (2009). *Effektiv planlegging og vurdering: Rubrikker og andre verktøy for lærere*. Bergen: Fagbokforlaget.

Harrow, A. (1972). *A taxonomy of psychomotor domain: a guide for developing behavioral objectives.* New York: David McKay.

Imsen, G. (2002). Lærerens verden. Innføring i generell didaktikk. Oslo: Universitetsforlaget.

Kirke-, utdannings- og forskningsdepartementet (2000-2001). *St.meld. nr. 27: Gjør din plikt - krev din rett.* Retrieved November 1st, 2009 from http://odin.dep.no/kd/norsk/dok/regpubl/stmeld/014001-040004/hov005-bn.html

Simpson, E. J. (1972). *The Classification of Educational Objectives in the Psychomotor Domain.* Washington, DC: Gryphon House.

Sirnes, S.M. (2005). "Flervalgsoppgaver – konstruksjon og analyse." Bergen: Fagbokforlaget.