Pedagogical use of multiple choice tests – Students create their own tests

Svend Andreas Horgen[1]

[1]*Sør-Trøndelag University College, Trondheim, Norway, svend.horgen@hist.no*

Formative use of multiple choice tests facilitates learning and can motivate and guide the student. The task of creating tests with questions and alternatives, can be difficult. The process towards constructing a successful test requires a certain amount of background knowledge, an overview of the particular field, and an ability to view the knowledge domain from a different perspective. This paper discusses an interesting way of using tests pedagogically: students are turned into producers, making their own tests. Theoretically, learning should be enhanced when students are exposed to such a creative task. Different experiments from real courses are discussed, and learning benefits and pitfalls discovered.

Keywords

Bloom's taxonomy, collaboration, pedagogical use of tests, quality assurance, student-production.

1. Introduction

Multiple choice tests can be used in assessment and to enhance the learning process. The straightforward way is for the teacher to develop multiple choice tests that students take. These tests can be used either in summative assessment, for instance as a substitute for a traditional exam, or in formative assessment, like self evaluation so that the student can test her own knowledge and be motivated to fill in knowledge gaps.

This paper presents an alternative way to use tests in order to enhance learning, motivated by work in the two projects "Pedagogical use of automated tests" (supported by the "Norway Opening Universities") [1] and an internal project called "UMIIK". Every teacher that has tried to make a multiple choice test, knows that this task can be truly challenging. Making a test of high quality is a process that requires knowledge and understanding of the field in question. Likewise, reflection, self-evaluation and quality assurance will be important processes in order to succeed. Such processes are also important for learning. Hence, implementing test creation as a student activity – where the students themselves create their own tests – should not only enhance learning, but also act as an interesting, beneficial pedagogical exercise for both the student and the teacher.

2. Theoretical foundation and motivation

According to Horton, making a good test is a process consisting of four iterative phases [2]. Several challenges exist in each phase. First, *test design* constitutes how to write good questions of high quality, but also involves deciding test properties such as feedback strategy. *Distribution* involves for instance how to deliver the test, the number of attempts allowed and how to deal with cheating. The phase of *grading* involves calculation of score and a grading strategy – tasks that are often both time consuming and difficult to do for the

average teacher – and should ideally be automated by using a test tool calculating correctly and efficiently. The final phase involves *improvement* of the test through analysis of the results so that for instance bad questions and alternatives can be identified. Improvement is an important part of the life cycle of a test, and typically would benefit from other people's comments and suggestions.

When creating a multiple choice test, the teacher meets challenges already in the first phase. How to write questions of high quality that not only test knowledge, but also test understanding, application and the ability to analyse a problem? Bloom's taxonomy [3] categorises cognitive levels of competence. It is quite possible to create multiple choice questions that test the student on different levels according to Bloom's taxonomy, thus making the test more demanding, valid and relevant. By providing the candidate with stimuli prior to a question, for instance an image, a sound/video clip, some text or other resources (or even activities), it is easier to test for instance the student's understanding, ability to apply knowledge or ability to analyse a problem. It is worth noting that the different cognitive levels are relative. What requires a deeper insight by a first-year student, may be perceived as routine recalling for the more experienced student [4]. Furthermore, how can a teacher create questions and alternatives that are not self-explanatory, and that allows the test to accurately depict the level of the student's knowledge and/or understanding? Identifying good distractors, that is, alternatives for a question that are incorrect, is a demanding task [5].

To summarise, for a teacher to succeed in making good questions, requires knowledge and understanding of the field, concentration, reflection, a critical eye, creativity and ideally an external evaluator for quality assurance.

Worth noting, a large question pool will be quite useful for the teacher when creating future tests and reusing existing questions. A pool can also assist in the tasks of result analysis and quality assurance in the improvement phase [6]. It obviously takes time to build a high quality question pool, but once populated, it would be very useful in order to create a variety of well designed tests more efficiently, for use in both formative and summative assessment.

3. Exercise: The students create their own tests

One of the main characteristics of the new successful Web 2.0-based services, is that of the *individual as the producer*. Traditionally in learning, the teacher is the one producing something, and the student is consuming the information. Producing something raises the level of participation, dedication and reflection, and is likely to result in learning. It typically involves new cognitive processing of existing knowledge. This applies to the production of written texts, presentations, learning content, etc., and should also be true for creating tests.

In order to succeed when creating a test consisting mainly of questions and alternatives, the test creator must view and process the theory and knowledge domain from a different angle than when performing tasks like traditional written exercises of explaining or reproducing something. For a multiple choice test, both detail-specific knowledge and a generic overview and understanding of the knowledge domain, is necessary in order to make successful questions.

Given that the teacher faces challenging tasks related to mental processing when creating multiple choice tests, the students should also meet those same issues. Hence – making a *pedagogical activity* out of *test and question production*, would seem interesting. In order to

succeed with their questions, students should have to read, process and understand the course material in a new, different way than before – and this activity should, in theory, enhance learning. Producing something obviously requires mental processing at different levels, so one might also relate the production activity itself to Bloom's taxonomy in general, and to the evaluation level (within Bloom's model) in particular. Self-evaluation of the test and question quality should also enrich learning, since reflection is regarded an important part of the overall learning process [7].

Last, but not the least, it is clear that alternative, creative activities can be fun. Salmon states the importance of creativity and fun "e-tivities" in order to enhance learning [8].

3.1 Individual exercise, spring 2006

During our courses, the students typically solve traditional exercises like "explain the reason for X" or "solve problem Y" or "create a web page using Z as a technology". Such exercises require the students to further process the theory after reading/attending lectures, and doing something practically.

As a new pedagogical learning experiment, the students in the courses "Internet Publishing" and "Operating systems with Linux" (spring 2006, different classes) were given a mandatory, individual exercise quite different to the ones they were used to, namely creating their own multiple choice tests. In addition to regular exercises, the students had previously answered multiple choice tests made by the teacher. Some questions were using images as stimuli. Obviously, such questions are more analytic in nature and require understanding rather than recalling facts in order to be answered properly. Therefore, all students had some basic understanding of the very nature of multiple choice tests and experience as test candidates.

Now chaired as producers, the students were challenged to do something, to them, totally new. In order to guide the process somewhat, a note explaining the importance of Bloom's taxonomy was attached to the exercise description, and the students were told to read this before creating their own tests. The introduction to the exercise listed well-defined, motivating learning goals: "Process the knowledge domain from a different perspective", "increase the ability to problematise knowledge", and "work with the cognitive levels of Bloom in mind". The exercise text also had a few examples of questions at different levels in Bloom's taxonomy. Finally, the last part of the exercise asked the students to explain/comment why their distractors were indeed wrong alternatives, and why their correct alternatives were correct – the idea being that question quality should improve further due to the reflective nature of this task.

When finished with the whole exercise, each student should evaluate its own learning process. Of course, it is impossible to conclude that the learning effect will increase within the bounds of such an activity – due to a small amount of empirical data – but student satisfaction and feedback from this particular exercise gives a certain indication of success. Almost every one of the about 30 students in each course enjoyed this form of exercise, and claimed to have learned more than they would in traditional exercises. Many students said that it was difficult to create good questions, but claimed the process to be fun and interesting. The fun-part was probably due to the fact that this activity was introduced about half-way through the semester, and regarded as a welcoming, new change compared with the traditional exercise form.

The work was intended to be individual, but some students still chose to collaborate (and admitted openly to have done so). Many questions were on the lowest level in Bloom's taxonomy. The students that collaborated had overall created better questions of a more analytic nature, thus succeeding in climbing Bloom's cognitive ladder. An interesting observation was that some questions and alternatives were really creatively formulated.

Technically, the students created their tests as Microsoft Word documents, which were read and assessed only by the teacher. Clearly, there was potential for improvement.

3.2 Collaborative exercise, spring 2007

Building on the experiences from the small experiment one year earlier, the students in the course "Operating systems with Linux" (spring 2007) were presented with a more thoroughly thought through set of integrated exercises. The overall goals were to ensure student participation, reflection, collaboration, motivation and variation.

The students were instructed to primarily work in groups of four. First, they were given an exercise where they should write some self-chosen texts into the course wiki. Worth noting, the wiki was also shared with two other related courses that the same students took simultaneously, thus making the wiki an important base of knowledge for the students. The wiki nature (public, shared, searchable, etc.) should encourage the students to better plan, work collaboratively and/or cooperatively, and not the least to reflect and process their articles with more attention to quality. The students should also pass a digital multiple choice test (individually) in the Linux course. The test was delivered through the LMS-system it's learning with a completion deadline and had questions aiming at testing understanding and application of knowledge.

Some weeks later, the students were given the exercise of creating their own multiple choice tests. Aiming at improving the pedagogical experiment from last year, the activity was now to happen within the well-established groups, and digitally. Technically, each student group was given *extended privileges* in their own folder in the LMS-system, so that it could effectively create and manage its own tests without intervention from other groups. The teacher had of course access to all group folders.

The students were given the same kind of introduction as last year, with a detailed description of the purpose and learning goals of the exercise, along with the theory of Bloom's taxonomy. The groups were instructed to produce one multiple choice test consisting of at least 10 questions. The students were encouraged to start by writing their own, individual tests, and then arrange a group meeting, solving, debating and discussing their questions. They should then agree on, and pick a subset of all the individual tests – a total of at least 10 questions, and fuse these into a single test representing the work of the group.

The intention with these guidelines was to ensure discussion and some degree of quality assurance within each group. These processes should enhance learning and stimulate activity on different levels according to Bloom's taxonomy. It was not explicitly stated in the exercise description that they should start to work individually and then pick a common subset, but a few groups still followed the oral encouragement and worked in such a way. This is shown in figure 1, where group 8 has created one common test (called "Flervalgsprove Team 8") consisting of a subset from the questions in the four individual tests. Note also that even though group 4 has just created one common test, its members

hopefully had collaborated and discussed the questions. In a worst case scenario, only one person has contributed.

Figure 1 Example of a group exercise where students have created their own tests within the LMS it's learning.

Each group were also instructed to write a reflection note evaluating its own work, the process, experiences, difficulties, degree and form of collaboration, subjection focus and perceived net learning effect. Each group were assessed based on both the question quality in their final test, and the corresponding reflection notes. The reflection notes were submitted using the assignment tool in the LMS-system, allowing the teacher to comment digitally on the work of each group.

The reflection notes (and the produced tests) showed that many groups had worked very well. Every group had fulfilled the minimum requirements, and most tests had questions from a broad range of the syllabus. Some questions were truly analytic in nature, aiming at testing the subject's understanding and ability to apply knowledge. Nobody used images as stimuli, although previous tests from the teacher had used images, and they were encouraged to do so. It would be interesting for later exercises to specify that at least a few questions should use images as stimuli, in order to enrichen the question quality. However, one should be careful not to loose focus on the real task, i.e. processing their knowledge on the field in a different way. It is important to note that some students in the 2007-experiment actually used code constructs in their stimuli, thus producing questions like:

"*given the command ls –la > res.txt on a directory with the 3 files a.txt, b.txt and c.txt, how many lines are then in the resulting file? Alternatives: 3, 5, zero*".

The correct answer is 5 (since the two hidden files . and .. are also processed due to the a-argument in the ls-command. This question requires the person being tested to have knowledge about the commands and analyse the specified command in order to answer correctly. The student who created such a question, clearly shows insight into how the recipient must think in order to solve the problem. Obviously, this is a hard exercise for a

regular student creating her first test (ever), and it is also a reconable explanation on why only a few students actually managed to climb Bloom's ladder in their produced tests.

Due to lack of functionality and flexibility in the LMS testing-tool, it would have been very time consuming to further analyse or do statistics upon the pool of student questions regarding for instance question quality.

Finally, after the deadline were passed and all the groups works were assessed, the teacher collected all the group tests and made them available for everyone to take. 8 group tests with between 10-20 questions, could theoretically yield a set of 80-160 different questions, but in practice, some questions are likely to either fully or partly overlap. The teacher copied all 8 tests to a special folder, and encouraged the students to take the tests. Such an activity would be interesting in more than one way. Obviously, the students get to test their own knowledge from a much larger set of questions than the teacher could have provided alone. Also, it should be interesting to see how other groups have focused and formulated their questions. Finally, as a producer of a test – knowing that co-students eventually will take the test – is likely to increase the dedication towards the task. Dedication and learning effect are closely related.

The course (along with the two closely related courses sharing the same wiki) had a multiple choice based exam halfway through the semester. At this point, the exercise of creating the group multiple choice tests, was submitted and assessed. The set of 8 tests should be a nice opportunity for every student to use as a preparation for the final exam. Reports from the LMS-system show that over 50% of the students tried out the test of the first and the test of the second group, but then the participants fall dramatically. Only five students completed the test of group 8. Perhaps the students found the questions to be too similar, or felt that the first two tests were sufficient in order to do self-assessment? A few students reported later on that they felt it was smart to do all the tests, but only had the energy to do a few.

Enhanced learning should be a primary goal of such an exercise. For the teacher, the pedagogical exercise presented in this paper is valuable in a practical way. There is likely to be produced a large set of questions, many of which would either be of high quality or could easily be edited to conform in such a manner. In the experiments of both 2006 and 2007, the teacher stated clearly that the produced questions would and should be used in tests for this course and coming courses. Hence, the students had another implicit encouragement to work well. Unfortunately, the testing-tool of the LMS "it's learning" is missing a pool feature and functionality to enable automatic fusion of many tests into one test. Some cut-and-paste would be necessary in lack of a better testing-tool in order for the teacher to reuse questions in different courses.

After finishing the course, the students were encouraged to evaluate the set of exercises through an anonymous survey. Unfortunately, only a third of the students responded, so the data set is fairly small. However, the results suggest that the activity of creating their own tests, was at least as useful as the wiki activity. Many students list the new, test activity as more fun. There were 6 exercises operating on the wiki (three interrelated courses, each with two wiki exercises), whereas there was only one exercise on "create your own multiple choice test". Thus – the least common exercise could benefit from the element of freshness. Whatever the reason, it seems clear that the alternative pedagogical activity was encouraging and motivating for the students. One student wrote "it is necessary to read quite a lot in order to be able to make good questions and alternatives".

4. Conclusions

Tests can be used formatively to enhance student motivation, activity and learning. As experienced by many teachers, there is a lot of learning and mental processing involved in the task of creating a test. It therefore makes sense to challenge students to create their own tests. This paper has explained in detail the various exercises given to the students, and discussed outcomes and learning effects. Experiments in a few courses have shown that many benefits can be obtained when implementing this kind of pedagogical activity as part of the pedagogical curriculum, regarding both enhanced learning and important testing aspects, for instance quality assurance, question pool and student acceptance.

References

1 http://norgesuniversitetet.no/seksjoner/english
2 Horton W. Designing Web-based training. Wiley 2000; 273-332.
3 Bloom. Bloom's taxonomy of Educational Objectives. 1956.
4 Brown G. Assessment: A Guide for Lecturers. November 2001. http://www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=3
5 Sirnes S. Flervalgstester, konstruksjon og analyse. Fagbokforlaget 2005.
6 Horgen S A, Larsen J, Hjertø G. Improving the Quality of Online Tests and Assessments. Educa Berlin 2005 Book of Abstracts 2005; 338-341.
7 SAPHE project. Self Assessment in Professional and Higher Education Project. http://www.swap.ac.uk/learning/assessment2.asp
8 Salmon G. e-tivities. RoutledgeFalmer 2002; 136-137.