Improving the Quality of Online Tests and Assessments

Svend Andreas Horgen, Jarle Larsen, Greta Hjertø
Sør-Trøndelag University College, HiST

Introduction

Online automated tests and assessment can play an important role in e-learning. The possibility of automated feedback and grading, as well as reuse of existing tests, increases the efficiency for the teacher. Tests provide more flexible learning environments and can also motivate and guide the student, due to immediate feedback.

However, it all depends on the quality of tests, and how they are used. Designing tests of high quality can be a demanding task for the teacher. Our work has focused on the development of a test tool that will help the teacher carry out automated testing. Using statistics in different ways, the tool will offer the teacher valuable means for analysing test results and improving tests.

Theoretical foundation, motivation and goals

We regard the skill of making good tests as an iterative process involving the following tasks: Design, Distribution, Grading and Improvement. [Horton2000]. The teacher face challenges in each phase.

Design: Most teachers don't have particular knowledge about how to write questions for use in automated tests. Best practices exists but can be difficult to apply in varying contexts without some failure or potential for improvement. Is it possible to make questions that reflect and test the real knowledge of the students? [Horton2000]

Distribution: There are many things to consider concerning test distribution. When should a test take place? How many tries should be available? How to ensure flexibility? How can cheat and technical problems be avoided? [Encheva+2005], [Buchanan1999].

Grading: Calculation of score is a non-trivial task. There are several ways to grade answers in automated tests. Reward of correct answers, punishment of wrong answers, avoiding alternatives that are mutual exclusive, etc. are some things the teacher must consider when choosing a grading strategy. [Bush2001].

Improvement: In order to improve a test, the previous phases need to be reconsidered in light of the results and experience from the test. The teacher needs to question what went wrong, what worked, the degree of difficulty, which questions were badly formulated, test properties, overall design, grading strategy etc. [Horton2000], [Encheva+2005].

Our goals are to achieve knowledge and develop methods in order to help the teacher through these phases, and find out to which extent tests can replace traditional assessment.

Hypothesis

Tests are an important part of the learning process. Tests can be used for assessment, self evaluation or be part of mandatory exercises in a learning program. Use of automated tests obviously releases valuable time for the teacher. We believe that online, automated tests can replace traditional assessments, *if* designed and implemented in a proper way.

The *quality* of a test depends on all the above mentioned phases, and implementing high quality tests clearly is an iterative process. Especially the last phase (of improvement) is important, but often neglected due to the amount of effort required. [Horton2000].

Evaluating the successfulness of a test can be done either with or without student participation. A retrospective evaluation form filled out by the students can reveal some mistakes, but is time-consuming both for the teacher and the students.

We believe that in most cases, tasks concerning improvement should happen without student involvement. Hence, in order to improve a test, the teacher has to revise and analyse the *test results*. A thorough analysis should include considering all answers and alternatives selected by the students. Mean (and other) values of scores, statistics of each reply, frequently chosen correct/wrong alternatives, completion time etc. are examples of valuable information for the teacher when revising questions and tests. For instance, if a wrong alternative to a multiple choice question was never selected by any student, it may signal that the alternative is badly written or "too obviously wrong". If all students with a high score all miss the same question, it might indicate the question or its alternatives being too diffuse or difficult to understand. Also, time-data is important. Information about mean, minimum and maximum completion times would certainly help the teacher in decisions like "there is too much time available on this test, I should tighten it from 30 to 20 minutes".

Implementation of a test tool

To support the life cycle of tests and improve test quality we have focused on developing a web based tool that records many aspects of tests that provide the grounds for test analysis.

The tool allows the teacher to create tests with questions ranging from multiple choice to free-text. To ensure flexible problem descriptions, a question can also refer to external resources, i.e. an image, web link or a movie clip. Furthermore, it is possible to write mathematical expressions in terms of Latex-code, and the tool will automatically produce the corresponding mathematical formulas as *images*, included among the rest of the text in the question. The same apply for everything else that can be expressed in Latex, i.e. physical and chemical formulas or music notes.

The teacher can control several important test parameters, like grading strategy, timing, availability, number of attempts, response strategy etc. To support import and export of tests, IMS QTI is implemented. [IMS]

All questions are saved in a searchable *question pool*. This facilitates reuse of existing questions. For instance, a teacher can set search parameters to search for those questions from the pool containing a specific word, belonging to a category, authored by a colleague etc.

The importance of improving tests and questions should not be underestimated. Our tool provides the teacher with useful *statistics* both at test level, question level and question alternative level. In this way, successful and unsuccessful questions and alternatives can be identified, thereby facilitating the analysis and revision tasks. Combining the pool with various statistics from the results of completed tests and question usage, the system can offer several interesting features that will assist the teacher in making tests of high quality with minimal effort required. The teacher might ask the system to generate a test with questions from the pool that meet certain criteria, and then edit this test further on. Other examples are easy development and deployment of test-templates and automatic generation of new questions based on existing questions that are considered successful. Some of these functionalities have not yet been implemented.

Finally, tests in such an online system are highly scalable. A well-designed tool is an important step in order to accomplish institutional goals concerning cost effectiveness, efficiency and flexibility for the student.

Use of the tool, practical research

We believe that the qualities of our tool will make the shift towards automated testing easier for the teacher. The tool is currently being used in real courses with both campus and distance learners. We are gaining experience and information both when students are using the tool sitting for an exam and through continuous testing during a course. The repository of test results is growing and is useful for the development of guidelines, and deriving knowledge about the phases of design, distribution, grading and improvement. Furthermore, experiences and wishes from users of the system, guide further development of the tool.

We are currently analysing test results, and comparing these with results from other, traditional tests, aiming at identifying conditions for automated tests to replace traditional assessment.

Conclusion and future work

Our work focuses on methods for developing tests, strategies for implementation and evaluation of tests, and assistance to the task of improving tests and questions.

We have developed a test tool to support our research and assist teachers. The searchable question pool facilitates reuse of tests between courses and teachers. Statistics help the teacher to analyse results and question success, which is important in the improvement phase. Other useful features, like automatic generation of questions, are yet to be implemented, but should even further support the life cycle of tests.

It is important to analyse the effects of such a tool. Though promising, we need to question whether it is helpful or not. Feedback from students could also reveal badly formulated questions not discovered by the system. Measuring quality of questions and tests are a difficult task, but a survey among the teachers using the system should reveal important aspects that will help improving the system and as such, help teachers implementing testing for assessment and as a pedagogical instrument.

References

[Buchanan1999] T. Buchanan: *Using the world wide web for formative assessment*. Journal of Educational Technology Systems, 27, 1998-1999.

[Bush2001] M. Bush: *A Multiple Choice Test that Rewards Partial Knowledge*. Journal of Further and Higher Education, Vol. 25 no 2, 2001.

[Encheva+2005] Sylvia Encheva, Sharil Tumin: *On multiple choice tests in a computer added assessment system*, 2005.

[Horton2000] W. Horton: *Designing web-based training*, Chapter 7. Wiley, 2000.

[IMS] World Wide Web, URL: http://www.imsglobal.org/