Assessing High-Level Thinking

SYLVIA ENCHEVA
Department Haugesund
Stord/Haugesund University College
Bjørnsonsg. 45, 5528 Haugesund
NORWAY
and
SHARIL TUMIN
IT Department
University of Bergen
P.O. Box 7800,
5020 Bergen
NORWAY

Abstract: - This paper focuses on Web-based assessment of students' conceptual thinking, information-processing skills and decision-making skills. Assessing the level of students' ability to use knowledge in an interrelated way when analyzing and solving problems is also discussed. This in contrast to traditional paper-and-pencil exam, evaluating students' knowledge of discrete components of a domain rather than the relations among these components. The following issues are discussed in particular: knowledge of different methods for solving a problem, abilities to recognize several correct answers and concepts, level of confidence in the correctness of their decision, ability to distinguish problems where the provided information is not enough for solving them (a `Not enough information is given' option) from problems which cannot be solved by a particular student only (a `Do not know' option). The impact of guessing is neutralized by combining level of confidence with negative marking.

Key-Words: - Web-based assessment, evaluation, level of confidence

1 Introduction

Research suggests that traditional methods of instruction produce greater success in domain specific areas [18], than those which assist students develop techniques which they can use in multiple topic areas. Most traditional methods of instruction do not seem to assess higher order cognitive skills like critical thinking and reasoning, and focus on discrete components of a domain rather than on the relations among these components.

There is a need for tests to assess the level of students' ability to use knowledge in an interrelated way when analyzing and solving authentic problems [14]. Questions in such tests will ask students to compare different methods, draw together several concepts, evaluate consequences and handle new situations. Understanding of important mathematical ideas and integrating more than one mathematical concept is discussed in [16], [17], [20] and [26].

Assessment of learning is a critical part of the learning and instruction process [22]. A set of criteria for alternative assessment practices has been formulated [1]. A method enabling the instructor to do a post-test correction to neutralize the impact of guessing is developed in [19].

More importantly multiple choice (MC) tests can be used to objectively assess cognitive, analytical and other comparatively high-level skills as well as straightforward factual knowledge [15]. More important, MC tests can provide impetus for improving instruction, and increase students' understanding of what they need to know and be able to do. Feedback helps to motivate students to learn and needs to be timely and constructive [4]. Students exhibit greater interest and levels of

learning when they are required to organize facts around major concepts and actively construct their own understanding of the concepts in a rich variety of contexts [24]. Research-based good practice addressing the pedagogical, operational, technological and strategic issues faced by those adopting computer assisted assessment is described in [2], [3], [8] and [21]. Effective question design is discussed in [25]. Expert and theoretical knowledge about the use of technology for assessment is offered in [6], [7] and [23].

In this paper we show how tests that use rules-based automatic scoring methods can assess some issues that traditional paper-and-pencil exam cannot: knowledge of different methods for solving a problem and abilities to recognize several correct answers and concepts level of confidence in the correctness of their decision combined with negative marking distinguish problems where the provided information is not enough for solving them (a `Not enough information is given' option) from problems that cannot be solved by a particular student only (a `Do not know' option).

The impact of guessing is neutralized by enforcement of negative marking and asking students to declare level of confidence for each answer they choose.

2 Various Test Types

A test item in an assessment contains *stems* (questions or incomplete sentences) followed by the corresponding putative answers or options - the key (correct option) and several *distracters* (incorrect options).

In some MC tests a student is asked to indicate her choice by clicking the radio button next to the answer, which implies that there is only one key among the putative answers. This way of grading does not reward possession of knowledge of different methods for solving a problem and abilities to recognize several correct answers and concepts. The traditional paper-and-pencil exam is not addressing these issues either.

Permutational multiple choice question tests have been used for assessing high-level thinking [13]. An item in such a test has two stems and six putative answers. Usually the student is asked to consider two similar concepts, or two complementary taxonomies. A question is answered correctly if each stem is matched up with the appropriate key. Again this implies only one correct answer. Such a grading does not distinguish a wrong answer caused by miscalculation from a wrong answer caused by lack of conceptual thinking or application of a wrong method.

Other MC tests consist of stems and putative answers where a correct answer gives full credit while an incorrect answer gives zero credit. This grading system does not differentiate between a choice of a wrong answer and choice of a `Don not know' option. Yet another type of MC tests applies negative marking for choosing a wrong answer to a particular question and thus improves on the problem with the previous type of MC tests. However, both ways of grading imply that a student is 100% confident in his/her answer.

Being able to properly judge the confidence of one's answers is an important part of being knowledgeable [15]. The paper describes experience from University College London where medical students are asked to state with each answer their level of confidence (1, 2 or 3) in the correctness of their decision. If the answer is correct, then this is the mark awarded and in a case of a wrong answer the corresponding mark is 0, -2 or -6. Such grading system implies only one correct answer.

One of the common pitfalls of traditional MC tests is that partial answers are not rewarded [12]. A form of MC tests rewarding partial answers called `liberal' is applied in [5]. A question there is followed by three options and a student may choose more than one answer. If the choice contains one option only, provided the correct one, he/she gets 100% for that question. If the student is uncertain and his/her choice consists of two or three options, again provided one of them is the correct one, he/she gets 67% or 33% respectively for that question.

2.1 Assessing High-Level Thinking

Students' conceptual thinking can be assessed by presenting them with tests where all the correct answers should be chosen and answers require integration of several components or approaches. A check-box may be used as a technical solution.

Consider a test where a student is asked to find all correct answers to every question in the test. The set of putative answers may contain several correct

answers and several wrong answers. There is no restriction on the number of correct and wrong options which prevents students from `intelligent' guessing about the number of correct answers following each question. In addition, we implement different factors for different options where the total sum of all factors after each question will be zero. Thus, if a student marks all of them, the resulting mark is zero. Application of different factors allows the examiner to reward partial knowledge and to enforce different deductions for cases like: a simple mistake in calculation, misunderstanding of a concept, lack of knowledge etc.

Assume that a particular question in a test is followed by five putative answers, where two of them are correct and the other three are wrong. The two correct answers have factors 0.6 and 0.4, while the three wrong answers have negative factors -0.1, -0.3 and -0.6 respectively. Calculating a mark for that particular question is shown in Table 1. If the final mark for that test is negative the system will give zero score.

Table 1: Calculation of a mark for a particular question

Putative answer	Factor	Student's choice	Mark (M)
a	0.4		
b	-0.6		
c	0.6	c	M = 0.6-0.3 = 0.3
d	-0.1		
e	-0.3	e	

Such a test is difficult to construct since the teacher should find good putative answers and should choose a suitable way to give the corresponding factors in case they are different. We assign higher factors to correct answers obtained by applying general methods and lower factors to correct answers requiring smart conceptual thinking. This way we are fair to the average student, recognize exceptional thinking and reward partial knowledge in a sense that a student may recognize only some of the correct answers and still get a positive mark. If the goal is to reward partial knowledge only, another type of assessment may be better applied. For wrong answers we assign the following factors: -0.1 for a small miscalculation that does not lead to a wrong answer to that question, -0.3 for a miscalculation that leads to a wrong answer and -0.6 for choosing a wrong method. Thus we differentiate between lack of understanding and a miscalculation.

Such tests should be constructed using inquiry-based learning principles. 'At the core of inquiry-based learning is the idea that complex problems may be vehicles for learning. Complex problems compel students to think about the many issues and alternatives inherent in the problems. This technique leads them to recognize that there may be multiple solutions' [28].

Such work should be done by a team of teachers. This saves time and helps to avoid unintended ambiguity which might happen since it is much more difficult for a teacher to write good multiple choice questions than to design essay questions.

3 How Much Confidence Do You Have in Your Answer

In our tests we enforce negative marking for a wrong choice and ask students to state a level (percent) of confidence (20%, 40%, 60%, 80% or 100%) after each answer. Thus a correct answer is awarded with a positive mark, a wrong answer by a corresponding negative mark and each of them is multiplied by a percent of confidence, selected by the student.

This way of marking is better than a traditional paper and pencil exam where a correct answer is assumed to be supported by 100% of the student's certainty and both a wrong answer and no answer have no effect on the final mark for the test. In that case is always beneficial for the student to write something, because it can improve the final grade if it is correct and does not make it worse if it is wrong. The Web-based assessment principals we use are related to the opinion that students should take part in the assessment process of their work. Another reason to involve this evaluation form in our practice is connected with real life situations. In their future working life students are supposed to make decisions and ask for expert's opinion in case of considerable doubt [27].

A `Not enough information is given' option is one of the putative answers following some stems. Students should experience MC tests where some questions may not contain enough information to answer the question or the given data in other questions is more than the information needed for

answering the question. Most text books contain only questions, examples and problems such that the provided data is enough to solve them. Further more all the given data must be used in the process of solving them. Usual exceptions are examples with systems of linear equations where the number of variables is more than the number of equations or the number of variables is fewer than the number of equations. Our experience indicates that most students do not apply concepts thought in one subject while working with another subject. For most of them `Not enough information is given' is equivalent to `Do not know'.

4 Rewarding Partial Answers

Details for rewarding partial answers are given in Table 2.

Table 2: Rewarding partial answers

Answer type	Factor
Correct answer	Full score
Correct information only but missing items	0.9
No missing items but including wrong items	0.5
Both missing items and containing wrong items	0.2
Wrong answer	-0.3

This form does not allow selection of more than one option. We find it more appropriate to allow a choice of several correct answers in another test type.

Remark: Automatic random ordering of answer types is applied while constructing tests. A decreasing factor ordering is used in Table 2 for the sake of presentation simplicity.

5 Experience With Applying Web-based Assessment

Method

We consider first year engineering students studying mathematics on undergraduate level. Traditional face-to-face teaching is supported by a Web-based asynchronous learning system. The system contains lecture notes, problems to be solved during tutorial hours, homework problems, formative and summative assessments. Material in lecture notes is divided in logically discrete instructional steps. Problem sections contain positive examples reinforcing understanding and negative examples establishing conceptual boundaries.

Activities are sequenced for increasing difficulty or complexity. To maximize learning efficiency, learners may be advised to repeat certain sections of material based on results from a diagnostic test, or tests within the sequence of learning activities. As a result of gradual removal of a tutor's support students became independent problem solvers.

An integral part of this work is a mechanism to provide instantaneous automated feedback to students. This encourages active involvement in the learning process, with the student taking more responsibility for his/her progress, and is thus an excellent preparation for the more self-motivated approach required. Prompt feedback is known to be related to student achievement and satisfaction [11].

In our grading system negative marking is applied within a question, i.e. negative marks are not carried over between questions. Negative marking means that some marks are deducted from overall test score for each wrong answer. Further more, different wrong answers are given different factors. Thus wrong responses are penalized by the degree of error.

Changing assessment methods encourages changing learning methods and results in students shifting from pure memorization to real learning [10]. We believe that different skills should be assessed by different test methods. Further more students are told explicitly about the scoring being used: either there is only one correct option among the putative answers or there are several correct options among the putative answers negative marking is either enforced or both wrong answered and not answered questions do not effect the final mark attach a confidence level or not reward partial answers or reward only correct answers.

If a student fails a test, he/she is provided with a diagnostic report showing his/her weaknesses. The student is advised to use the diagnostic report to improve both his/her learning and performance for future tests.

6 Student's Perceptions of Web-Based Tutorials and Web-Based Assessment

At the end of each course students were asked to indicate on a five-point Likert scale how true statements about usefulness of Web-based tutorials were for them. The data indicate that 95% of those students (N=276) found the materials useful for understanding of the subject.

Investigation of student perceptions of the Web-based self assessments revealed that the students were using them as a learning tool as well as for the original purpose.

6.1 Scoring Distribution

Course grades for two groups of undergraduate engineering students in a calculus course are analyzed. The control group had no access to Web-based tutorials and Web-based assessments. The experimental group had access to Web-based tutorials and Web-based assessments. The control group contained 86 students enrolled in the course during Fall semester 2004. The experimental group contained 93 students enrolled in the course during Fall semester 2005.

The results for the experimental group were consistent with their performance in the course, where those who scored high overall results in the final exam also scored high marks for their Web-based tests. Applying Web-based assessments resulted in a wider distribution of scores, indicating that such grading helped the rates in the group make more discriminating judgments about the quality of the presentations.

Conclusion

To prepare students for a job situation where they have to deal with problems that are not yet recognized, they should experience solving complex problems with multiple correct answers and requirement for making a decision in a limited amount of time.

In our tests we enforce negative marking for a wrong choice and include an option button where a student declares a level of confidence in the correctness of each answer. Thus a correct answer is awarded with a positive mark, a wrong answer by a corresponding negative mark and both are multiplied by a level of confidence, declared by the student. This way of marking is better than a traditional paper and pencil exam where a correct answer is assumed to be supported by 100% of the student's certainty and both a wrong answer and no answer have no effect on the final mark. In that case is always beneficial for the student to write something, because it can improve the final grade if it is correct and does not make it worse if it is wrong.

We believe that different skills should be assessed by different test methods. Further more students should be told explicitly about the scoring being used - either there is only one correct option or there are several correct options among the putative answers, negative marking is either enforced or both wrong answered and not answered question do not effect the final mark, attach a confidence level or not, reward partial answers or reward only correct answers.

The formal evaluation elicited useful perceptions concerning the effectiveness of the way in which Web-based materials and assessments were integrated into lectures and tutorials. Recommendations arising from this formal evaluation will be used for redesigning features in the next version of these materials and assessments.

References:

[1] M. Birenbaum, Assessment 2000: towards a pluralistic approach to assessment, *Alternatives in Assessment of Achievement, Learning Process and Prior Knowledge*, Boston: Kluwer, 1996.

[2] J. Bull and S. Zakrzewski, The mass implementation and evaluation of computer-based assessments. *Assessment and Evaluation in Higher Education*, Vol. 23, No. 2, pp.141-152, 1998.

[3] J. Bull and D. Stephens, The use of question mark software for formative and summative assessment in two universities, *Innovations in Education and Training International*, Vol. 36, No. 2, pp. 128-136, 1999.

[4] J. Bull and C. McKenna, *Blueprint for Computer- assisted Assessment*, London: Rutledge, 2003.

[5] M. Bush, A multiple choice test that rewards

partial knowledge, *Journal of Further and Higher Education*, Vol. 25, No. 2, pp. 157-163, 2001.

[6] G. Conole, and A. Oliver, A pedagogical framework for embedding IT into the curriculum, *The Association for Learning Technology Journal*, Vol. 6, No. 2, pp. 4-16, 1998.

[7] G. Conole, E. Crewe, M. Oliver and J. Harvey, A toolkit for supporting evaluation, *The Association for Learning Technology Journal*, Vol. 9, No.1, pp. 38-49, 2001.

[8] P. Deubel, An investigation of behaviorist and cognitive approaches to instructional multimedia design. *Journal of Educational Multimedia and Hypermedia*, Vol.12, No. 1, pp. 63-90, 2003.

[9] F. Dochy and L. McDowell, Assessment as tool for learning. *Studies in Educational Evaluation*, Vol. 23, No. 4, pp. 279-298, 1997.

[10] F. Dochy and G. Moerkerke, The present, the past and the future of achievement testing and performance assessment. *International Journal of Educational Research*, Vol. 27, No. 5, pp. 415-432, 1997.

[11] Dunkin, M. J. and Barnes, J. *Handbook of Research on Teaching*. Hanover: American Educational Research Association, 1997.

[12] J. Engelbrecht and A. Harding, Online assessment in mathematics: Multiple assessment formats, *New Zealand Journal of Mathematics*, Vol. 32, pp. 57-66, 1998.

[13] D. W. Farthing, Permutational multiple choice questions: An objective and efficient alternative to essay-type examination questions, *Proceedings of 3rd Annual Conference on Integrating Technology into Computer Science Education* (ITiCSE'98), Dublin, pp. 81-85, 1998.

[14] R. Custer , B. G. Valesey and B. N. Burke An assessment model for a design approach to technological problem solving, *Journal of Thechnology Education*, Vol. 12, No. 2, 2001.

[15] A. R. Gardner-Medwin, Confidence assessment in the teaching of basic science. *Association for Learning Technology Journal*, Vol. 3, pp. 80-85, 1995.

[16] G. Gibbs, S. Habeshaw and T. Habeshaw, *53 Interesting ways to assess your students*, 2nd ed., Bristol: Technical and education services, 1988.

[17] A. H. Johnstone and A. Ambusaidi, Fixed-response questions with a difference. *Chemistry education: Research and practice in Europe*, Vol. 2, No. 3, pp.313-327, 2001.

[18] T. de Jong, and M. G. M. Ferguson-Hessler, Types and qualities of knowledge. *Educational Psychologist*, Vol. 31, pp. 105-113, 1996.

[19] R. Harper, Correcting computer- based assessments for guessing, *Journal of Computer Assisted Learning*, Vol. 19, pp. 2-8, 2003.

[20] D. Lawson, Formative assessment using computer-aided assessment. *Teaching mathematics and its applications*, Vol. 18, No. 4, pp.155-158, 1999.

[21] G. Moerkerke, Assessment for flexible learning. *Lemma te Utrecht*, Open Universiteit Nederland, 1996.

[22] D. J. Nicol and D. Macfarlane-Dick, Formative Assessment and self-regulated learning: a model and seven principles of good feedback practice, *Studies in Higher Education*, Vol. 31, No 2, pp.199-218, 2006.

[23] M. Oliver, J. MacBean,, G. Conole, and J. Harvey, Using a toolkit to support the evaluation of learning. *Journal of Computer Assisted Learning*, Vol. 18, No. 2, 199-208, 2002.

[24] G. Polia, *How to Solve It*, NJ: Princeton University Press, 1957.

[25] N. Pritchett, Effective Question Design', *Computer Assisted Assessment in Higher Education*, London: Kogan, pp. 29-47, 1999.

[26] G. Smith, L. Wood, M. Coupland, B. Stephenson, K. Crawfordand and G. Ball, Constructing mathematical examinations to assess a range of knowledge and skills, *International journal for mathematics education in science and technology, Vol.* 27, No. 1, pp. 65-77, 1996.

[27] T. O. Williams, Jr., A._M. Fall, R. C. Eaves, and S. Woods-Groves, The reliability of scores for the draw-a-person intellectual ability test for children, adolescents, and adults, *Journal of Psychoeducational Assessment*, Vol. 24 pp. 137-144, 2006.

[28] S. D. Thompson, L. Martin, L. Richards, and D. Branson, Assessing critical thinking and problem solving using a Web-base curriculum for students, *Internet and Higher Education*, Vol. 6, pp. 185-191, 2003.