Assessment based on exercise work and multiple-choice tests

JAN KETIL RØD, SVEINUNG EIKSUND & OLAV FJÆR
Department of Geography, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT Higher education institutions in Norway have implemented reform as part of the Bologna Agreement. The reform also aims to follow up students more closely and to change the way in which they learn by increasing the amount of formative assessment, including grading. This paper presents a case study of a first-year methodology course in geography, in which students’ performance in exercise work and tests based on multiple-choice questions was graded. Multiple-choice tests were used to increase the level of reflection during exercise work and to encourage students to read continuously throughout the semester.

KEY WORDS: Assessment, multiple choice, exercise work, learning geographic methodology

Introduction
A comprehensive reform of higher education, the Quality Reform, was implemented in all higher education institutions in autumn 2003 in Norway. It represents Norway’s adoption of the Bologna Agreement by standardizing summarizing assessments and academic degrees and thereby facilitating increased mobility of students and faculty members within Europe. The reform also aims to increase study intensity, internationalization and the quality of higher education. One suggestion leading to an increase in both the number of credits earned by students each year (study intensity) and student performance (study quality) was to reduce the amount of summative assessment and instead increase the use of formative assessment (St.Meld, 2001).

Formative assessment has long been part of geography courses on methodology and geospatial technologies such as GIS and remote sensing. In these subjects, there has always been a strong component of compulsory exercise work intended to provide students with skills in the collection, analysis and presentation of geographical data, often with the use of computer software. Students hand in their results and receive feedback from their tutors. Other courses not focusing on methodologies or technologies have traditionally been taught by lectures with a lesser amount (or no amount) of student activity involving formative assessment. In Norway, however, there was an increase in the use of portfolios based on term papers in many geography courses after the reform, but it
created unfortunate opportunities for cheating, for instance, through plagiarism in their essays, as students could copy partly or wholly from each other or from articles found on the Internet. Assessments of plagiarized work are not valid and do not provide tutors with information on how well students have mastered a given subject. Dysthe (2007) therefore advises the use of counter-strategies: educating students generally in the use of sources and particularly Internet sources, as well as associated ethical issues, and to use anti-plagiarism software. A more serious problem for many higher educational institutions in Norway, however, is the cost involved in implementing the Quality Reform, with its many formative assessments.

This paper provides a brief overview of an attempt to implement continuous formative and formal assessment in a first-year introductory course in quantitative methods for geographers held at the Department of Geography at NTNU. It is a 7.5 credit (European Credit Transfer System) course and its content is descriptive statistics, GIS and remote sensing. The course is rather practical in nature, with a considerable amount of formative assessment based on feedback on exercise work which, since 2008, we use as formal assessment. We grade students’ performance on exercises and related multiple-choice questions (MCQs). Here, the term ‘multiple choice’ is not only used for items with three or more answer options, but also true/false items and matching questions (paired lists that require students to correctly identify the relationship between the items). The combined results from six MCQs and six practical exercises account for 20 per cent of the final grade (exercises 12 per cent and MCQs 8 per cent), whereas the end-of-course examination influences the final grade by 80 per cent. Doing practical exercise work, ticking the correct answers in MCQs and writing short essays as part of the end-of-course exams represent assessments of qualitatively different aspects of cognition. The general issue in focus in this paper, then, is how we should design a framework for accommodating formal assessments for the different aspects of cognition within rigid economic frames.

**Assessment and Cognitive Taxonomy**

Gold *et al.* (1991, p. 136) define assessment as “the process which measures educational attainment on the basis of student performance and which provides feedback to the student about the strengths and weakness of that performance”. A variety of assessment methods are available, and a general principle forwarded by Gold *et al.* (1991), which we adopt here, is to assess what is valued and to reward good performance. Although the Bloom cognitive taxonomy (Bloom *et al.*, 1956) is still highly influential for educators who need to define, teach and assess student achievements, we will use a simpler cognitive classification system for the various levels of scholastics aptitude that we appreciate and therefore should assess for the particular methodology course under study. We employ Haladyna’s (2004) three aspects of cognition – knowledge, skill and ability – that resemble the revised taxonomy of educational objectives (Anderson & Krathwohl, 2001). The first entry, knowledge, is a cognitive process that Haladyna (2004) divides into two levels: recalling and understanding (Table 1).

Although several authors claim that multiple-choice assessments are also applicable for higher levels in Bloom’s cognitive taxonomy (Haladyna, 2004; Sirnes, 2005), we consider that they may only be useful for assessing the lowest parts, namely recalling and understanding.
Assessment based on the End-of-Course Exam

Since the implementation of the Quality Reform, there has been increased use of e-portfolios and other forms of assessments at higher education institution in Norway, but these alternative forms of assessment have rarely replaced the traditional end-of-course exam (Dysthe, 2007). One exception is distance learning courses, where e-portfolios are a convenient form of assessment. For on-campus courses, it is more common to find alternative assessments in addition to, and not as a replacement for, end-of-course exams.

The aim of the end-of-course exam is to assess a student’s ability regarding problem-solving and critical thinking (see Table 1). For instance, in the end-of-course exam in 2008 one of the tasks showed a two-column table, where each row represented a country in Europe as well as its urban population percentage. Students were asked to classify the variable (urban population) and suggest how the classed variable could be presented in a map. Although this is a practical problem to solve, an outstanding solution would require a critical discussion of the various issues of thematic mapping and the associated implications.

Assessment Based on Compulsory Exercise Work

National and international literature on students’ perceptions of and study strategies towards compulsory work often portray students negatively. Gibbs (2006) claims that students are increasingly strategic in the way they allocate their time and effort and may only study what is assessed. Dysthe (2007) finds a similar negative pattern for Norwegian students: students do the minimum that it takes to have their work approved, but not more. Such less-motivated students may invest only a minimal amount of effort in the compulsory work, but make an all-out effort on the days just before their exam. They may therefore succeed in their exam results but soon forget what they have learned. However, when knowledge is gained rapidly, often it is also lost rapidly (Gynnild, 2003).

Our previous experience in teaching this methodology course has shown that there are students who invest only the minimum effort in the compulsory parts. On the other hand, there are also many students who input a huge amount of effort in doing the exercises and as a result obtain a high level of procedural knowledge or skills in the use of statistical packages and GIS. However, for the end-of-course exam, which demands maturity regarding students’ cognitive ability, not all students are able to build up to peak performance for this event, with the unfortunate consequence that their efforts do not ‘pay off’ when their final grades are assessed. Both of these outcomes are unfortunate and should be avoided.

Through seven exercises (six of them marked), students had practice in managing and presenting geographical data in various ways using both a statistical software...
package (SPSS) and geographical information systems (ArcGIS). As recognized by several researchers and stated by Gibbs, “skills are learned through cycles of practice and feedback” (Gibbs, 2006, p. 19), but the value of feedback is reduced if it comes too late. For courses with many students, it is common for feedback to be slow, particularly if the feedback is based on essays or term papers. However, as the sheer volume of commenting on exercise work is moderate and can be done more rapidly, our assessment was normally ready after three days. It is our hope that three days were sufficiently prompt for this assessment to have been formative, and thus to have supported subsequent learning during the course of study for this particular subject (Lambert, 1997).

**Assessment based on Multiple Choice Questions (MCQs)**

According to Hogg, “MCQs tests do have their limitation and should not be used as the only means of assessing students” (Hogg, 1997, p. 125), but as we allow students’ performance on MCQs to influence only 8 per cent of their final grade, we do not consider this warning to be applicable. We had two reasons for wanting to include MCQs in this course. Our primary goal was to increase students’ reflections on what they do during the compulsory exercise work, and a secondary aim was to stimulate continuous reading throughout the semester and not only in the final days before the exam. Since the results of MCQs influence students’ final grades, we believe that few students will use the study strategy of relying on peak performance in the days immediately prior to their final exam.

Table 2 presents three examples of multiple-choice questions used for statistics, cartography/GIS and remote sensing. The latter example demonstrates the use of matching questions.

In total, we conducted six tests with 120 questions, approximately 20 per cent of them based on matching alternatives, 70 per cent based on multiple-choice questions (three or more categories), and 10 per cent with two categories (true/false questions). The six MCQs were all computer-delivered through our e-learning system. All MCQs were administered in an open-book situation and since the MCQ tests had no time limit students could spend as much time as they wanted to in order to find an answer to the question from textbooks, articles or on the Internet. This is contrary to recommended practice, where tests with ample time for completion are regarded as reducing the tests’ reliability (Burton, 2005). Nevertheless, the allowance of ample time to complete a test corresponds with our second objective: to function as an encouragement to start reading early and to read more rather than to be an objective test of students’ knowledge.

Although technically easy, it is difficult to design test questions that stimulate students’ interest in reading and reflecting on what they read. Sirnes (2005) gives some guidelines on how to design good multiple-choice questions, such as that they should not be transcripts from textbooks. Hence, we did our best not to formulate questions in such a way that students would be able to copy their answers directly from a text, but instead would need to reflect on an issue or perform some calculations (as demonstrated by question 1 in Table 2) in order to select the correct answer. Consequently, whatever sources the students had available would not be of use to them if they did not understand the basic principles involved.
Five Issues to be Discussed

As already mentioned, our general objective for this paper is how to design a framework for accommodating formal and formative assessments for different aspects of cognition. In particular, we are interested in how to assess students’ practical knowledge gained from exercise work and also their capabilities of recalling and understanding what they read. We are also interested in how we can couple the MCQs with exercise work in order to increase students’ learning outcome from exercise work. We address these issues by posing five questions:

**Issue 1:** How to increase students’ learning outcome from exercises?

**Issue 2:** How to make exercises less open to plagiarism and hence more legitimate for formal assessment?

**Issue 3:** How to design individual MCQs applicable for both formative and formal assessment?

**Issue 4:** How to formulate the multiple-choice questions and alternatives in order to stimulate reasoning rather than guessing?

**Issue 5:** How to establish how students react to having to do the exercises and take the tests?
Discussion

Issue 1: How to increase students’ learning outcome from exercises?

All exercises are based on the use of a statistical package (SPSS) and a GIS (ArcGIS), and we have written the exercise texts as step-by-step instructions. Continuous assessment of exercise work is, according to Gold et al. (1991), an appropriate assessment of practical knowledge or skills, but our previous experience compares more closely to the following description: “[students] might simply following instructions to press certain buttons without even thinking why they need to do so. This is surface learning, rather than deep learning” (Scheyvens et al., 2008, p. 60).

One alternative could be to let students find out for themselves how to operate the software, but since these software packages are rather complex for first-year students their user interfaces should be customized to the tasks which students need to perform (Lloyd, 2001). Another alternative that we tried was to add several questions between the steps, thereby breaking up the streamlined form of the exercise. The questions were either some factual issues for the student to find an answer to or some theoretical issues for them to reflect on. The questions were repeated in an MCQ test.

Many lecturers maintain that MCQs “are not good at testing for high-level cognitive learning” (Nicol & Milligan, 2006, p. 67), whereas for students the use of MCQ tests alone will not provide any good learning outcome for those faced with the task of finding the correct answer (key) rather than engaging in interpretation and thinking (Paxton, 2000, p. 111). Other argue that MCQs “can be designed to go beyond factual recall and to measure students’ abilities in intellectual reasoning, in problem solving and in numeracy” (Chalkley, 1997, p. 102). The usefulness of MCQs, however, will depend on how tests are constructed and how they are coupled with other forms of learning activity. Hogg (1997) refers to an approach of coupling MCQs with lectures, where the questions in the test relate to the topics being taught in the lecture. The argument he uses is that since “the lecture continues to be a major vehicle of instruction in most universities, the need to make sure that it is stimulating and effective in communicating key concepts, theories and methods remains of paramount importance” (Hogg, 1997, p. 124). We use a similar argument regarding exercise work which continues to be used as a vehicle for training in quantitative geographical methodologies. Thus, it is important that the time students spend on exercise work provides deep learning.

Issue 2: How to make exercises less open to plagiarism and hence more legitimate for formal assessment?

Exercise work can be open to plagiarism. Students may simply copy the whole or parts of an exercise from a fellow student or from a student from a previous year. Since plagiarism is common in subjects with compulsory exercise hand-ins (Franklyn-Stokes & Newstead, 1995), we were determined to design exercises in a way that made cheating difficult. Although we strongly discourage plagiarism, we equally strongly encourage cooperation between students by giving them the same tasks. Although there are some instances of assessing collaborative work by students (e.g. Stanier, 1997; Webb, 1997), there is a strong tendency in higher education in Norway to perform individual student assessments. We followed the common practice and produced individual exercises by dividing huge datasets thematically. We used 27 variables describing the countries of the world by...
socioeconomic characteristics, downloaded from a World Bank website (http://devdata.worldbank.org/dataonline/). All variables were from the year 2005 and we assigned a unique pair of variables to each student, for which they had to perform a variety of tasks including univariate and bivariate descriptive statistics and mapping. There were 78 students following the course in 2008 and all of them had a different combination of variables. For instance, student ‘A’ had to work with variables on GDP per capita and life expectancy at birth for females whereas student ‘B’ had to work with variables on GDP per capita and life expectancy at birth for males. Students thus had to solve the exercise tasks by using their own data and as a result they could not simply copy the results from a fellow student. Consequently, we achieved a framework for applying formative as well as formal assessment to students’ exercise work, but without applying strict examination conditions.

**Issue 3: How to design individual MCQs applicable for both formative and formal assessment?**

A problem relating to plagiarism exists also for MCQs; namely, that students should be assessed individually. In order to hinder students from copying answers from fellow students, ideally we either had to force students to take the test simultaneously or have unique questions for each student. The former option was not appealing for two reasons: it would have required us to put the students under a form of control similar to an exam situation, and it did not offer the students any flexibility. With many students, the latter option was not appealing either, as that would have required us to formulate a considerable number of questions. Instead, we made a compromise. For each of the six tests, we designed a pool of 20 questions and each student had to answer five questions randomly selected from the pool. By calculating the binominal coefficient, this yields that for each test 15,504 different samples of questions are possible. The probability that two fellow students would have all, four, three, two, one, or no questions in common equals the number of favourable outcomes divided by the number of possible outcomes. These probabilities are presented in Table 3.

The probability that two fellow students would have exactly the same five questions is very low, \( p(\text{Five equal}) = 0.00006 \), although there is a 0.44 probability that they would have one question in common. We consider these probabilities to represent an acceptably low level of cheating possibilities.

As we had five MCQs, each student was posed 30 questions in total. Although some people may “have faith in tests with even fewer than 60 items, even as few as 30” (Burton, 2005, p. 70), we agree with Burton that “60 items are too few for adequate sampling of a large knowledge domain” (2005, p. 70). However, as we do not intend to use MCQs to test the complete knowledge domain, but only to stimulate an increased level of reflection during exercise work as well as continuous reading, we believe 30 items to be sufficient.

**Issue 4: How to formulate the multiple-choice questions and alternatives in order to stimulate reasoning rather than guessing**

We used the e-learning system It’s Learning to design the multiple-choice tests and after deadlines were passed the system automatically generated assessment marks. In addition to generating assessment marks and how many points a student had obtained, the e-learning system reported whether the answers were correct or not, thus providing
formative feedback. Although Nicol’s mapping of MCQs to the seven principles of good feedback practice states under principle six – closing the gap between desired and actual performance – that “students can repeatedly take MCQs and check answers until they reach a satisfactory performance” (Nicol, 2007, p. 56), we only allowed each student to take a test once. In the case of another course where we tried out the use of MCQs, we let each student have three attempts at answering the MCQs. We experienced that allowing more attempts on MCQs neither reduced the amount of guessing nor increased effort in terms of continuous reading.

Nevertheless, also with only one attempt students may guess the correct answer. “Any test taker when encountering the item answer either knows the right answer, has partial knowledge that allows for the elimination of implausible distractions and guess among the remaining choices, or simply guesses in the absence of any knowledge” (Haladyna, 2004, p. 217). Also important is the fact that guesses are not blind, as they may not be completely uninformed and random: “Guessing will be less often blind if poorly constructed test items contain unintended clues, such as bad wording (e.g. grammatical or syntactical clues) or obviously implausible distractors” (Burton, 2005, p. 67). We therefore scrupulously wrote questions and alternatives in order to avoid any clues to correct alternatives or flagging of unlikely alternatives. However, what would be most useful in order to stimulate reasoning rather than guessing is probably what Bush (2001) suggests, namely to “reward examinees who possess partial knowledge as compared with those who are merely guessing” (Bush, 2001, p. 157). Bush (2001) suggests using negative marking, whereas Gardner-Medwin (2006) provides an example of how additional reflection could be obtained when students provide a confidence rating alongside their MCQ response. According to this method, in addition to marking correct answers, students must rate their degree of certainty on a three-point scale. Both their answers and their confidence level influence their score on the MCQ; correct answers are

<table>
<thead>
<tr>
<th></th>
<th># favourable</th>
<th># possible</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Five equal)</td>
<td>(\binom{5}{5} \binom{15}{0} = 1)</td>
<td>(\binom{20}{5} = 15504)</td>
<td>0.00006</td>
</tr>
<tr>
<td>P(Four equal)</td>
<td>(\binom{5}{4} \binom{15}{1} = 75)</td>
<td>(\binom{20}{5} = 15504)</td>
<td>0.00484</td>
</tr>
<tr>
<td>P(Three equal)</td>
<td>(\binom{5}{3} \binom{15}{2} = 1050)</td>
<td>(\binom{20}{5} = 15504)</td>
<td>0.06772</td>
</tr>
<tr>
<td>P(Two equal)</td>
<td>(\binom{5}{2} \binom{15}{3} = 4550)</td>
<td>(\binom{20}{5} = 15504)</td>
<td>0.29347</td>
</tr>
<tr>
<td>P(One equal)</td>
<td>(\binom{5}{1} \binom{15}{4} = 6825)</td>
<td>(\binom{20}{5} = 15504)</td>
<td>0.44021</td>
</tr>
<tr>
<td>P(Zero equal)</td>
<td>(\binom{5}{0} \binom{15}{5} = 3003)</td>
<td>(\binom{20}{5} = 15504)</td>
<td>0.19369</td>
</tr>
</tbody>
</table>
weighted positively by the confidence level but incorrect answers are weighted negatively. However, as It’s Learning does not provide any collection of students’ confidence levels, we did not use confidence-based marking.

**Issue 5: How to establish how students react to having to do the exercises and take the tests?**

According to Hovdhaugen et al. (2007) it is often the students’ own apprehension of what they have gained from their studies which is the most realistic method for measuring learning outcome. Towards the end of the spring semester 2008, we therefore designed an electronic questionnaire survey asking students to evaluate the course and their learning outcome. Of the 78 students that followed the course in 2008, 48 responded. In total, we asked them 22 questions, which were in both open-ended and closed format. Seven of the questions and student responses are presented in Table 4. Possible responses ranged from 1 to 6, where 6 was the highest rating. Overall, the majority of students rated their satisfaction with the course as four or higher (question 1, Table 4).

We asked the course participants to rate their own learning outcome from exercises (question 2, Table 4) and MCQ tests (question 3, Table 4). The majority of students reported a rate of 4 or higher for their learning outcome, the percentage being much higher for the exercises (79.2 per cent) than for the MCQs (58.3 per cent). We also asked them three questions with the aim of identifying whether or not the results of exercise work and MCQ tests made them more committed. The large majority (89.5 per cent) reported that they had made an extra effort with the exercises (question 3, Table 4) and a high majority (75–79.2 per cent) reported that due to the MCQ tests they read more both of the course literature (question 5, Table 4) and of their lecture notes (question 6, Table 4) than they would have done otherwise.

There will always be an element of guessing in MCQ tests. In our case study, students who found the questions too tricky, and were unable to reason which alternative represented the correct answer, made a guess:

> When one could not find anything about the question anywhere, one had to guess.

> A couple of times I had questions that I was unable to find the answers to from the syllabus, even if I searched for a long time. Then I guessed and probably got the wrong answer! That is boring....

> I guessed the answers to a few of the questions that I found unclear or where I was unable to find a firm answer in the syllabus.

To some extent, all of the students must have experienced uncertainty as to the correct answer since only 8.3 per cent claimed that they never guessed the answers (see question 7, Table 4). However, the third open response listed above was from a student among the 8.3 per cent of students who responded that they never guessed (question 6, Table 4). There may be other inconsistencies in the results of our questionnaire survey and thus also an underreported percentage of guesses. We recognize also that many of the closed questions may have been leading as they solicited a ‘Yes’ or ‘No’ response. A lesson we have learned for the future is to reformulate questions as statements and let the students...
Table 4. Students’ rating from course evaluation (NA = Not answered)

<table>
<thead>
<tr>
<th>Responses (%)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 How satisfied are you, in general, with the course? (1 = not satisfied, 6 = very satisfied)</td>
<td>0</td>
<td>4.2</td>
<td>8.3</td>
<td>33.3</td>
<td>37.5</td>
<td>16.7</td>
<td></td>
</tr>
<tr>
<td>2 How would you evaluate your learning outcome from the exercises? (1 = very poor, 6 = very good)</td>
<td>2.1</td>
<td>2.1</td>
<td>16.7</td>
<td>45.8</td>
<td>18.8</td>
<td>14.6</td>
<td></td>
</tr>
<tr>
<td>3 Did you make an extra effort with the exercises knowing that the results would influence the final grade? (1 = not at all, 6 = yes indeed)</td>
<td>2.1</td>
<td>2.1</td>
<td>6.3</td>
<td>20.8</td>
<td>33.3</td>
<td>35.4</td>
<td></td>
</tr>
<tr>
<td>4 How would you evaluate your learning outcome from the test questions? (1 = very poor, 6 = very good)</td>
<td>2.1</td>
<td>12.5</td>
<td>25</td>
<td>29.2</td>
<td>20.8</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td>5 Did you make an extra effort reading the course literature knowing that your results from MCQ tests would influence your final grade? (1 = not at all, 6 = yes indeed)</td>
<td>2.1</td>
<td>6.3</td>
<td>14.6</td>
<td>25</td>
<td>22.9</td>
<td>27.1</td>
<td>2.1</td>
</tr>
<tr>
<td>6 Did having to answer test questions make you work more with the lecture notes than you would have done otherwise? (1 = not at all, 6 = yes indeed)</td>
<td>0</td>
<td>10.4</td>
<td>10.4</td>
<td>41.7</td>
<td>16.7</td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td>7 As the test questions were in a multiple-choice form, did you guess the answers? (1 = always, 6 = never)</td>
<td>2.1</td>
<td>2.1</td>
<td>10.4</td>
<td>22.9</td>
<td>52.1</td>
<td>8.3</td>
<td>2.1</td>
</tr>
</tbody>
</table>
indicate their strength of agreement or disagreement by matching their responses to categories on a Likert scale.

Although we did not particularly ask students for their views regarding the fact that results from exercises and MCQ tests influenced their final grade, we nevertheless received many responses to the open-ended questions, some of which are given as follows:

I think that the multiple-choice questions forced me in a positive way to read during the academic year.

It is a very fine measure that exercises influence the final grade.

I think I learned a lot and I think it is great that the results from exercises count as a percentage of the final mark.

Exercises are an excellent type of work. They give me a change to work with the compulsory reading little by little. I learned a lot from the exercises.

There were, of course, also some negative sentiments expressed about the exercises and MCQs, but none of them relating to the fact that they influenced the final grade.

**Concluding Remarks**

Given the diversity of students at Norwegian universities, in common with universities elsewhere, the validity of using only a single form of assessment is jeopardized. In order for assessment to be equitable, multiple and varied forms of assessment should be used that allow for students who are disadvantaged in one type of assessment to have the opportunity to prove their knowledge, skills and abilities in other types of assessment (Gipps & Murphy, 1994).

In this paper we have discussed our experience in assessing student performance using three types of assessment in a first-year geography course. In addition to the traditional end-of course exam, we assessed student results from compulsory exercise work as well as from multiple-choice questions. Although the results obtained from the three types of assessment correlated (bivariate Pearson’s r yield 0.41, 0.51 and 0.59) and we were able to predict the exam results based on scores from exercises and MCQ quite closely (regression analysis yielded an adjusted R square 0.42), some of the residuals were rather high. Put differently, although there was overlap among the 10 best results from exam, exercises and MCQs, there were also some pronounced differences. As seen from results from the questionnaire survey, overall students were content with the course and very few had reason to be dissatisfied with the arrangement of diversified assessment. Although most of them would have received the same grade without the influencing exercises and MCQs, 14 students (18 per cent) had their marks improved whereas only two (3 per cent) had their marks lowered.

According to Gibbs (1999) appropriate assessment engages students in exactly the kind of learning activity that is desirable. If we want to change the nature of how students read, the best strategy could be to require students to write essays: “To write an essay you need to ‘read around’ a topic to develop an argument… Assessment substitutes such as multiple-choice questions are extremely unlikely to generate this kind of reading”
We have not used multiple-choice questions as a substitutional assessment but as an additional assessment. We wanted to change the nature of how students performed the compulsory exercise work, in order to increase their level of reflection while doing the exercises. By rewarding well-performed exercise as well as well-answered multiple-choice test where questions posed in the exercise text are repeated, it pays for students to put in extra effort. Student feedback informed us that they had learned a lot from the exercise work, and hence this seems to be a good strategy.

With a high amount of exercise work and a huge number of multiple-choice tests, there is a risk of continuous formative assessment having a dampening effect on students’ enthusiasm for study (Gold et al., 1991). However, as the majority of students in our case study, in general, were satisfied with the course, we believe that the size of the workload we gave to our students did not have any dampening effect at all. Normally, it is not having too much formative assessment that is problematic, but rather that it comes too late, is too general, vague, without guidance, focuses on the negative parts, or is unrelated to assessment criteria (Weaver, 2006). In order to help students to recognize quality, it may be important to involve them in both the development of assessment criteria and performing assessments. When teaching the methodology course in 2009 we have therefore experimented with self-assessment and may try out peer assessment in 2010.

Acknowledgements

The authors would like to thank Catriona Turner for her useful advice and four anonymous reviewers for their very useful comments.

References


