Effect of self-assessment on test scores: student perceptions

Beatriz U. Ramirez
Faculty of Medical Sciences, Universidad de Santiago de Chile, Santiago, Chile

Submitted 2 February 2010; accepted in final form 11 May 2010

Ramirez BU. Effect of self-assessment on test scores: student perceptions. Adv Physiol Educ 34: 134–136, 2010—After a sudden increase in most of the individual grades in a multiple-choice test, students were asked to rank the three most relevant factors responsible for this outcome. Among eight others, the availability of a test for self-assessment before the final test was by far the most frequently mentioned (82.4% of the students). Questions applied during different course activities did not have the same effect on student scores as the “online” self-assessment test.

Methods

A group of 47 second-year undergraduated students took my physiology course and, among other assessments, received three multiple-choice tests (MCQ) throughout the semester. The course has a webpage (moodle platform) with supporting material (PowerPoint slides used in lectures, cases, problems, etc.) accessible to all students. Moodle is an open-source, completely free to use course management system (9) provided to staff and students in our university.

Questions during class time were applied during the whole course. The first MCQ test was printed and applied in the usual classroom. In the second and third MCQ tests, the students were present but the test was applied online using the facilities of the moodle platform in a building named CITECAMP, where each student accessed a computer. Otherwise, the three tests were similar.

Before the second test, but not before the first or third test, a MCQ test given the previous year that covered the same topics as second test was posted online for student self-assessment and training (they had no previous experience with online tests). Students could access the self-assessment test anytime for a 1-wk period up to 8 h before the second test was given. They were allowed to repeat the test three times and instantly received the corresponding scores, which were valid only for their own self-assessment process. Scores were percentages of correct answers.

One week after the second test, students were asked to voluntary answer a survey intended to rank the factors that could have contributed to their increased success in the second test. Students were requested to list the main factors that affected their improvement. They could list as many factors as they wanted. Only the three most relevant factors were to be ranked.

One of the forty-seven students did not attend the three tests and was eliminated from the study. Informed written consent was given by all students, and there was no incentive for participation.

Data were analyzed using Dunnet’s multiple-comparison test after ANOVA. Statistical significance was set at \( P < 0.05 \).

Results

The mean scores of the three tests are shown in Table 1. Only 5 of 46 students did not get a higher score in the second test. The increase ranked from 0.7% to 33.4% of correct answers (14.05 ± 0.08, mean increase ± SD), and in the second test, half of the students got >67% correct answers compared with the former 57% (Table 1).

A total of 34 students answered the survey and ranked the factors that contributed to their success. The five students that did not increase their marks in the second test did not respond to the survey (by their own choice). A list was made with all factors that got the three higher ranks individually; these were considered the “main factors.” In total, there were nine main factors. The frequency of each of the factors was then measured. Figure 1 shows the impact of these factors, expressed as the percentage of students that considered each factor to be one of the three most relevant. As shown in Fig. 1, students considered the ability to test their knowledge (online self-assessment test) before the second test to be the most relevant
factor (82.4% of the students). All other factors were by far less relevant. Lecture attendance was considered a relevant factor mainly by the few students that were not attending them before the first test. To be nervous during the first test was considered a secondary factor by a few students, but no one gave this factor a high priority. The difference between a printed and an online test was not mentioned as a factor affecting student outcome.

With these results in mind, from then on, multiple choice-type questions were more frequently applied orally (sometimes the question was written in a slide) in every session (interacting lectures or small-group seminars for problem solving or case analysis) to reinforce student self-assessments of their knowledge and reasoning capacity. A student had to answer the question, and he/she could discuss the answer and the underlying explanations for a few minutes with their classmates (those sitting close) before giving the final answer. Each answer was followed by an open discussion among all students (Was the answer right/wrong? Was the explanation incomplete? Why?).

Then, it was time for the third test. Despite the fact that there was an increase in the number of questions done during class time, student marks were lower than in the second test and not statistically different from the first test (Table 1).

**DISCUSSION**

MCQs are widely used for measuring knowledge, comprehension, and the application of learning outcomes (4). According to historical data, the three MCQ tests applied here were of similar difficulty, but the first test was printed and given in the usual classroom and the second and third tests were applied online, in the more comfortable CITECAMP room. Despite the fact that these differences may affect student perceptions of the quality of teaching, they were not considered by them to be relevant to their outcome.

Student performance was significantly better in the second test than in the other two tests. However, the second and third tests were given online in the same room. In the three tests, questions had a similar degree of difficulty, a large amount of supporting material was available, and questions covering the subject were applied and discussed during class time. Therefore, the main difference between the three tests was that the self-assessment process was applied only in class in the first and third tests and both in class and online in the second test. It appears that students learn better when they have the chance to self-assess their knowledge at their own pace.

Learning improves when the learner is mentally actively engaged in the process of learning (1, 10). Therefore, in lectures and seminars, opportunities were given to predict the expected outcomes and to verbalize and discuss the answers to correct misconceptions (7). Also, students were encouraged to reason causally and qualitatively about physiological situations to improve their critical-thinking skills and foster learning with understanding (8). However, only a few students volunteered to participate in these discussions, and it was not possible to assess if the others were involved in the learning process. Most students were not open to engaging themselves in discussion in front of their classmates. In contrast, all students had access to the self-assessment test posted online.

In addition, since the online MCQ test for self-assessment contained questions used the former year to assess the same topic, there was a better match between the question style and test style in the self-assessment test than in the questions used during class time. Moreover, the online self-assessment test allowed three repeats; therefore, feedback was often provided. Thus, students could direct their study by focusing on the aspects needed to strengthen their weakest knowledge in their own way (talking alone or with others, reading, writing, drawing, or other) and at their own pace. This probably provided them with the conditions needed to understand what they were learning and, thus, to reach more meaningful learning, that is, long-lasting and transferable learning that develops the ability to extend what has been learned in one context to new contexts (1, 2, 8).

The amount of time taken to learn is variable among subjects (1), and the in class discussion time may have not been enough for those who needed more time to master the subjects. In addition, tests can directly affect learning by promoting better retention, due to an increase in the amount of study time, to an improvement in study strategy, or both (5).

Among the nine main factors mentioned by the students, the self-assessment test was by far the most relevant, since it was mentioned by >80% of them, with the closest other factor mentioned by <40% of the students. Other factors, like a comfortable place (quiet), more study, or availability of supporting material, were considered relevant by 33–38% of the students, and stress was considered a minor issue. All these factors may affect learning, but the susceptibility to them is highly variable among subjects (1).

The decrease in the outcome of the third test was an unexpected result and strongly suggests that the feedback

### Table 1. Results of the three multiple-choice question tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Percentage of Correct Answers</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td>56.7 ± 10.3</td>
<td>56.7</td>
</tr>
<tr>
<td>Test 2</td>
<td>68.5 ± 9.7</td>
<td>67.4</td>
</tr>
<tr>
<td>Test 3</td>
<td>60.1 ± 9.4</td>
<td>59.2</td>
</tr>
</tbody>
</table>

Values for percentages of correct answers are means ± SD. All students completed the three tests. Statistical analysis of the individual data indicated that test 2 was different from test 1 and that test 3 was similar to test 1 (by Dunnet’s multiple-comparison test after ANOVA: \( P < 0.05 \), test 1 vs. test 2; and \( P = \) not significant, test 1 vs. test 3).

---

![Fig. 1. Factors that affect grade improvement: student self-assessment. The open survey question was as follows: “Rank the three most relevant factors in the improvement of your grade in the test 2 compared with test 1.” CITECAMP is a comfortable room used in test 2 but not in test 1.](http://advan.physiology.org/)

---
provided by the self-assessment test was not replaced by an increase in the in class feedback.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES