A progressive assessment strategy improves student learning and perceived course quality in undergraduate physiology

D. A. Saint, D. Horton, A. Yool, and A. Elliott

Discipline of Physiology, School of Medical Sciences, University of Adelaide, Adelaide, South Australia, Australia

Submitted 15 January 2015; accepted in final form 26 May 2015

Saint DA, Horton D, Yool A, Elliott A. A progressive assessment strategy improves student learning and perceived course quality in undergraduate physiology. Adv Physiol Educ 39: 218–222, 2015; doi:10.1152/advan.00004.2015.—In 2010, second-year physiology (n = 165) had a traditional single 3-h end-of-semester exam. To provide diagnostic feedback earlier, for students enrolled in 2011 (n = 128), we incorporated an in-class exam at 3 wk in addition to the final exam. Based on initial analysis and positive student comments, for the 2012 cohort (n = 148), we expanded this to incorporate four 1-h in-class exams every 3 wk plus a short final integrative exam. Average scores from exams and questionnaires (student evaluations of learning and teaching, 10 questions) were compared among 2010, 2011, and 2012 cohorts. We also compared scores in the practical component of the course, which had a constant format for all cohorts. Data are given as means ± SD; statistical analyses were done with unpaired two-way Students t-tests. From 2010 to 2012, there was a significant improvement in total exam scores (59.7 ± 15.8 vs. 68.6 ± 14.2, P < 0.001) but no significant change in total practical scores (72.3 ± 9.0 vs. 74.4 ± 10.2, P = 0.05), indicating that the rise in exam score was not due to higher academic abilities of the 2012 cohort. Overall mean student evaluation of learning and teaching responses (4.9 ± 0.4 vs. 5.3 ± 0.3, P = 0.015) and overall percent broad agreement (66.0 ± 8.0 vs. 79.2 ± 7.5, P = 0.003) indicated a significant improvement in student satisfaction. In conclusion, both learning outcome and perceived course quality were enhanced by the increased frequency of summative assessment, possibly by promoting consistent student study habits.

IT HAS BECOME AXIOMATIC that “assessment drives learning,” but there is still some debate about whether this is positive or negative (6, 23). Some negative aspects may be that, depending on the type of assessment, assessment can stimulate surface rather than deep learning, or “learning for the test” (22). The way in which assessment is integrated into the overall curriculum can also produce negative effects on student learning (17). For example, a large end-of-course test designed to assess primarily recall may drive students to “cram,” or study intensively in the short period before the test, rather than studying consistently throughout the course (14, 27). The result can be that students are less engaged with the early part of the course and, hence, will have poorer recall of the earlier material compared with that presented later (9). One way commonly used to address this type of problem is to incorporate formative assessment early in the course (24). Formative assessment is useful if structured feedback is provided promptly after the assessment and if the material is retested later in an overlapping summative assessment. In other words, formative assessment is only useful if the student has the opportunity to act upon the results (33). However, a less tractable problem with purely formative assessment (i.e., assessment of material that is not reassessed later) is that since the assessment does not count toward final grades, one is relying on the motivation of the student to fully commit effort to the assessment task and to take any corrective action necessary. This is not always the case (26, 29), perhaps because students lack the necessary “meta-learning” skills (3) or because they perceive the task to have little reward (31).

To address these problems, we have adopted a strategy in which assessment could function as formative but that is still graded and counts for a proportion of final grades, a “hybrid” formative-summative assessment, spread throughout the course. It has been shown that a structure such as this, in which assessment is spread in time through the course, can improve retention (20), especially if material is repeated in the different tests.

An important principle of such “progressive” or “cumulative” assessment is that it can function as a powerful formative assessment only if the students have the opportunity to compensate for poor performance in an early assessment by an improved score in subsequent tests. In other words, each individual assessment must not contribute so much to the eventual overall mark that it dominates the overall assessment; otherwise, poor performance in an early assessment will act as a demotivator (“why bother, I’m going to fail anyway”). On the other hand, each individual assessment must not contribute so little as to be trivial; otherwise, in a way similar to purely formative assessments, the students may exert little effort [a low “task value” in the expectancy-value model of motivation (31)] and, hence, be unable to correctly calibrate their learning progress.

Hence, a well-designed assessment strategy incorporates a compensatory element (students can correct poor initial performance by subsequent improvement), but, similarly, students cannot rely overly on a good initial performance and so reduce their effort later. A well-designed assessment strategy should therefore result in individual students that initially perform poorly improving their later performance, but, at the same time, students that initially perform well are encouraged to maintain this high standard.

In the present study, we report the results of a modification of the assessment strategy in a second-year physiology course that we implemented over 2 yr. Before the changes, the course (running over 12 wk) assessed learning in a single 3-h exam held about 2 wk after the end of classes. For the cohort of students enrolled in the next year, we incorporated a short exam 3 wk into the course, to allow students to assess whether they were learning at the rate and level expected. Positive feedback from the students and improved student scores after this change prompted us to go further and to spread the
Assessment throughout the course for the students taking the course in the following year (four 1-h exams at 3-wk intervals, testing each “module” of the course, followed by an end-of-semester online integrative exam that covered all the material). For the 3 yr, the same material was presented and examined at the same level.

This change was well received by the students, as indicated by feedback on their student evaluation of learning and teaching (SELT) and resulted in an overall improvement of exam scores.

METHODS

Bachelor of Health Science students at the University of Adelaide take physiology in the second and third years of the program. Teaching in second-year physiology includes lectures on four modules (sensory, motor control, cardiovascular, and respiratory physiology), taught by three lecturers over a 12-wk semester as well as online tutorials and a practical component. Lecture notes are made available before the lectures, and lectures are recorded and made available online. Lecture attendance is not compulsory. In 2010 (n = 165) and in previous years, the course was assessed by a final 3-h exam, which counted for 50% of the total mark, with the balance made up of tutorials (15%) and the practical component (35%).

Aiming to encourage progressive learning and to provide diagnostic feedback earlier, in 2011 (a different cohort of students, n = 128), we incorporated a 1-h in-class exam at 3 wk in addition to the final exam (10% for the 3-wk exam and 40% for the final). Based on the initial analysis and positive student comments, in 2012 (n = 148 for this cohort), we incorporated four 1-h in-class exams at 3-wk intervals plus a short integrative final exam after the end of the semester, with each exam counting for 10%.

At the end of the semester, students were asked to complete a SELT, which asked about their experience of the course in the form of 10 questions answered on a 7-point Likert scale and also asked them to write a short comment on those aspects that they liked and those that they disliked.

Average scores from exams and SELT questionnaires were compared among 2010, 2011, and 2012 cohorts. We also compared scores in the practical component of the course, which had a constant format for all cohorts. Data are given as means ± SD; n values are numbers of students. Statistical analyses were done with unpaired two-way Students t-tests.

RESULTS

Effect on learning. In 2010, the overall mean mark in the exam for the course was 59.7 ± 16.4 (mean ± SD) and the proportion of students failing the exam component (<50% score) was 27% (45 of 165 students). After the change implemented in 2011 in which we introduced a hybrid assessment in week 3, the overall mean mark was not significantly changed (58.6 ± 14.5) with the proportion of students failing essentially unchanged (31%, 40 of 128 students). A correlation analysis for this year showed that students who performed poorly in the formative assessment task (exam A) tended to improve their performance in the final exam (exam B). In contrast, higher-performing students in exam A showed a slight reduction in performance during the final exam (r = 0.78, P < 0.001; Fig. 1, the dashed line is a slope of 1 and the solid line is regression of the data).

After the changes implemented in 2012, with progressive assessment throughout the semester, the overall exam mark improved significantly (68.6 ± 14.2, P < 0.001 compared with 2011; Fig. 2) and the proportion of the class failing fell to 12% (18 of 148 students).

Feedback from students. We analyzed responses from SELT surveys. Students were surveyed in 2010 and 2012 using the same method, which consisted of a standard set of 10 questions about all aspects of the course answered using a 7-point Likert scale. The results are shown in Fig. 3 as mean responses to each of the questions in 2010 and 2012 or the percent broad agreement in those years for each question. This latter measure is the proportion of the students giving an answer of 5 or more on the Likert scale for that question.
Both overall mean and percent broad agreement increased in 2012 compared with 2010 (mean score: 4.9 ± 0.4 in 2010 vs. 5.3 ± 0.3 in 2012, P = 0.015, and overall percent broad agreement: 66.0 ± 8.0 in 2010 vs. 79.2 ± 7.5 in 2012, P = 0.003). Of the 10 questions asked, question 6 (“Uses methods of assessment that help achieve learning outcomes”) was directly related to assessment. The answer to question 6 showed a particularly striking improvement (2012: mean score of 5.4 ± 1.4 and percent broad agreement of 78% vs. 2010: mean score of 4.4 ± 2.6 and percent broad agreement of 56%). The other questions in the survey as well as responses to them are shown in Fig. 3.

**Qualitative feedback.** In the 2012 SELT surveys, there were 120 responses. As part of the SELT survey, we asked students to comment on the following question: “What are the best aspects of this course, and why.” Fifty-nine of one hundred twenty of the comments mentioned the new assessment strategy. Example comments are as follows:

Exams after each topic, though stressful, were good as they tested knowledge ‘as it came.’

Multiple ‘exams’ assist in promoting study throughout the semester.

The exams being 10% each, not 3 hours exam at end of semester worth 70%.

We also asked the students to comment on the following statement: “This course could be changed in the following ways to improve my learning.” Only 2 of 120 students commented negatively on the exam format, as follows:

Prefer one final exam as it would be easier to study for to gain the best possible mark...

 Didn’t like the exam format.

Internal control. Since we were testing different student cohorts in each year, the improvement in outcome may have been due to differences between the cohorts. However, since the practical component did not change between the years, it acted as an internal control. The difference in cohort scores in the practical between 2010 and 2012 was not statistically significant (72.3 ± 0.7 vs. 74.4 ± 0.8, P = 0.05, confidence interval: −4.22 to 0.03), which we interpret as indicating that there was no intrinsic difference in ability between student cohorts.

**DISCUSSION**

**Problems with the traditional assessment process.** Assessing learning largely by a single end-of-semester exam (50% of final mark for the course) obviates the possibility of students assessing their learning progress throughout the semester and deprives them of the opportunity to correct poor progress before the final summative assessment. Although students could compensate to some extent by a good performance in the tutorials (15% of final grade) and the practical component of the course (35% of the final grade), these were aimed at different learning skills, such as group work, and not necessarily aligned with the major course content in the lectures.

Having a single summative assessment also had negative effects on student strategies for learning. Lecture attendance is not compulsory in the course, and although attendance was not logged, we have previously noted that it is generally poor. This is likely to be because of the availability of the recorded online material, extra notes, and other sources of content. Although we have previously shown that poor lecture attendance is not necessarily reflected in poor performance (13), poor lecture attendance may signal a lack of engagement with the material. This was evidenced by monitoring the timing of downloads of the online material; most occurred in the week before the exam, after the teaching semester had ended.

**Stage 1: addition of an early formative-summative assessment in 2011.** Assessment is only formative if it leads to action either by the teacher or student (5). To achieve any outcome, it has to incorporate feedback (7). With this in mind, after the assessment task at 3 wk in 2011, we promptly gave students...
their individual marks and a breakdown of the marks for the class. Although there was some overlap between the early formative assessment and the end-of-semester exam, we did not specifically retest content in the final exam. Rather, the feedback provided was aimed at allowing students to assess whether they were learning in sufficient breadth and depth (19). In other words, we wanted students to assess their learning, not as “am I rote learning enough of the material?” but rather “am I learning the material in sufficient depth, and doing sufficient extra reading, to do well in the type of questions being asked?”. As a consequence of the formative assessment and feedback provided, we assumed that poorly performing students would “lift their game” and improve their performance. Superficially, this may have occurred: as shown in Fig. 1, students that did poorly in exam A did, in general, improve their score in exam B.

If the addition of early formative assessment was effective in producing better learning outcomes, this should be reflected in the overall scores of the students in that year compared with scores in the course without the formative-summative assessment. This was not apparent in the results between 2010 and 2011, suggesting that adding this early assessment task did not improve overall scores (with the caveat that we measured different cohorts of students and so cannot accurately assess any potential small effects). This inconsistency in the effect of early formative assessment has been shown in a large cohort of medical students (15); the authors attributed the inconsistency to other confounding factors, such as student perception of imminence, impact, and workload. Similarly, a retrospective analysis of assessment in a cohort of dental students found no correlation between scores in formative assessments and final grades (1). The type of formative assessment may make a difference; a study of 146 medical students in the United States showed that an open-book untimed formative assessment had a larger effect on learning than did a closed-book timed assessment (18). In contrast, a study of a 548 biomedical students across a group of Spanish universities showed unequivocally that the addition of a midterm formative assessment improved student performance in the final summative assessment (8). Interestingly, the authors noted that participation was a better predictor of final outcome than success in the formative assessment, a study that included voluntary formative assessments spread throughout an undergraduate medical course (9 assessments in all) showed that participation in the formative assessment was a powerful predictor of end-of-course marks (30). Interestingly, in a way similar to that found for single formative assessments (18), the type of online formative assessment appears to influence the outcome. For example, of four online quizzes given to students in an introductory physiology course, only one quiz was associated with an increase in exam performance (21). Indeed, it has been suggested that formative assessment can have detrimental effects on learning (28), especially if it is in the form of multiple-choice questions with no feedback, since this results in “false learning” of the distractors in the questions.

To minimize these confounding factors, we made the assessments compulsory (in the sense that they incorporated enough marks to have an impact on final grades if students elected not to take them) and provided prompt feedback: students had access to the correct answers after each exam. This change to an assessment strategy spread throughout the semester did improve student scores as reflected in overall marks for the course. We considered the possibility that this improvement may have been due to this being a different cohort of students, but the existence of an internal control in the form of the practical assessments, which did not change between the years, provided reassurance that the abilities of the different cohorts were broadly similar. Given this and the finding that the change to a “progressive” assessment strategy received very positive feedback from students, we conclude that this strategy encourages student engagement with the course throughout the semester and leads to better learning outcomes. However, this improvement in learning outcomes has to be balanced against the extra time and effort in administering the extra assessment tasks.

Conclusions. There is still not a clear consensus in the literature about the effectiveness or otherwise of formative assessment. In an early wide ranging landmark study (4), Black and Williams provided a review of the literature on formative assessment in which they concluded that formative assessment produces improvements in student learning with effect sizes between 0.4 and 0.7. However, the authors themselves pointed out that the diversity of studies included meant that overall conclusions were difficult. Indeed, subsequent literature has not widely supported the view that formative assessment improves learning outcomes, at least in K–12 students (16, 32). The literature is similarly inconclusive at the tertiary education
level (33), although there seems to be a growing consensus that online formative assessment can be effective (12).

In the present study, we have shown that, at a university level, both student learning outcomes and student satisfaction with a second-year physiology course were enhanced by an increased frequency of examinations throughout the course (a progressive assessment strategy) and associated feedback.

REFERENCES

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

AUTHOR CONTRIBUTIONS

Author contributions: D.A.S., D.M.H., A.Y., and A.E. conception and design of research; D.A.S. and D.M.H., A.Y., and A.E. performed experiments; D.A.S., D.M.H., A.Y., and A.E. analyzed data; D.A.S., D.M.H., A.Y., and A.E. interpreted results of experiments; D.A.S. and A.E. drafted manuscript; D.A.S., D.M.H., A.Y., and A.E. edited and revised manuscript; D.A.S., A.Y., and A.E. approved final version of manuscript; A.E. prepared figures.

RECENTS