Assessment: can it lead to a better course design?

Journal of College Science Teaching, September-October 2013

Author abstract

Concerns about the misuse of assessment results in accountability-related efforts led our biology faculty to take the initiative to develop our own assessment tool for one of our introductory biology courses. Our goal was to evaluate student achievement in the course and detect methods for improving our course. We developed a pre/posttest, which we used for two semesters. Results from these two semesters showed a very small but significant increase from pre-to posttest scores. For the third semester of the study, we moved to a collegewide common final. By creating questions based on the learning outcomes of the course, we were able to compare scores for specific questions to learning outcomes. These comparisons provided insight into the difficulty of certain concepts in biology, along with the degree of uniformity of instruction across various sections of the course. These results led us to question the amount of information covered in the introductory course, the difficulty level of the course, and the vague nature of our learning outcomes. Presently, on the basis of the results, our biology faculty are in the process of reevaluating the content of the course and developing new, more detailed course learning outcomes.

Assessment has always been an integral part of college teaching, and teachers have traditionally used varying types of assessment to determine how much students learn. However, because of political pressure, accreditation demands, and economic concerns, there has been a growing trend to combine assessment with accountability (Harris, 2011; Liu, 2011; Snyder & Carnicom, 2011; Tompkins & Cates, 2009; Volkwein, 2010). Currently here in Florida, there is the additional concern that our governor is leaning toward the Texas model of faculty accountability (June, 2011; Mangan, 2010, 2011), in which assessment results were used to award bonuses, promotions, or tenure.

A controversial new book, Academically Adrift: Limited Learning on College Campuses (Arum & Roksa, 2011), has raised some serious questions regarding the ability of a wide range of American colleges to provide a satisfactory level of education. The authors point out that this controversy has been heightened by an absence of research on the topic.

These concerns led us to begin a serious attempt to develop an objective method to evaluate the effectiveness of our introductory biology course. We wanted to be able to develop a collegewide tool that would not only allow us to evaluate students across four campuses, but also help us to improve our course. In several recent studies (Elfrink, Kirkpatrick, Nininger, & Schubert, 2010; Mohanty, Gretes, Flowers, Algozzine, & Spooner, 2005; Parish & Allen, 2011; Siebert, Siebert, & Spaulding-Givens, 2006), assessment has been based on a learning-centered approach using pre/posttests to evaluate course outcomes. The use of learning outcomes (LOs) to design quality assessment tools has also been well documented (O'Dell, 2009; Posner, 2011). This not only enables the instructor to design effective assessment tools, but also helps ensure that the test questions are directly related to the topics covered (Posner, 2011). We decided to use this method for the current study. We were then able to evaluate the effectiveness of the overall course and each LO separately.

Methods

At Edison State College, we offer an Introduction to Biological Sciences (BSC1005) course for students who may not have had sufficient high school biology to enroll into Biological Science I (BSC1010). BSC1005 is a three-credit course that has no lab component. We chose this course because it has a large enrollment, is offered on four different campuses, and is taught by different instructors who are required to use a common syllabus with specific LOs.

A multiple-choice test covering all of the 12 LOs in the course syllabus was developed by faculty who teach the course. As a starting point for developing the exam, some of the questions were based on a standard database of questions that came with the required textbook; others were created by faculty who teach the course. Although time constraints and expense prevented us from validating our exam, our faculty worked together to continuously edit and revise the questions, which ensured that our exam questions not only reflected the LOs, but also were consistent with accepted canons for creating reliable and suitable test questions (Haladyna, Downing, & Rodriguez, 2002; Popham, 2009). The exam was used as a pre/posttest for two consecutive fall semesters (2010, 2011). At the beginning of the spring 2012 semester, a departmental decision was made to switch from a pre/posttest to a common final.

At the beginning of our fall 2010 semester, each course instructor was given a copy of the test with instructions to administer it immediately. Students were asked to write their name and student ID number on each answer sheet. The tests were machine-graded on a Scantron and the results for each individual were downloaded into an Excel spreadsheet. The exact test was given again at the end of the semester, and the results were downloaded again. For purposes of the study, only the results from students who took both the pre- and posttests and had correctly entered their student IDs were used in the data analysis; in total, data from 211 students were used.

Because we discovered that, in some instances, the pretest was not administered until 2 or 3 weeks after the start of the course, and the posttest was given 2 to 3 weeks before the course was over, the procedure was modified for the fall 2011 semester. In order to ensure LOs were not covered in the course before the pretest and after the posttest were administered, instructors were asked to administer the pretest within the first week of class and the posttest during finals week. Again, only results from students who took both the pre- and posttests and had correctly entered their student IDs were used in the study, for a total of 191 students. For spring 2012 the answers to the questions on the original exam were scrambled and the exam was used as a common final. In total, data from 387 students were used in this study.

Results and discussion

Small increase from pre- to posttest scores

The most significant finding from the pre/postest study is the small but statistically significant increase in the average pre/posttest scores for both fall 2010 and fall 2011 (Table 1). Although the scores on the pre- and posttest were statistically different ($p < .01$), it was a surprise that the average difference between pre- and postscores was so small (8% in 2010 and 13% in 2011). Not only were the differences between pre/posttest scores similar for the two semesters, the scores for the posttest were also similar--53% for fall 2010 and 52% for fall 2011.

These data suggest that little increase in learning took place in the introductory biology course. We feel that there may have been two major factors that contributed to this finding. First, students may not have taken the test seriously because the results did not impact their final grade. Instructors voiced concern that the exams took up valuable class time, and their attitude that the exams were a nuisance may have been passed on to their students. Second, even though the importance of following the timeline was made clear to all instructors, the tests may not have been given in most classes according to the required timeline.

As a result, we decided to abandon the pre/postexam for the spring 2012 semester and use the exam simply as a "common final" in all sections of the course collegewide. This resulted in students and instructors treating the exam more seriously, and all exams were given during finals week. This did produce a moderate increase in scores; the average score on the common final in spring 2012 was 61% (Table 2).

Differences in posttest scores among the course LOs

By creating our exam questions based on course LOs, we were not only able to gain insight on the structure and content of the course, but also on the structure and content of our exam. Table 3 is a list of the 12 LOs as listed on our common syllabus.

Figure 1 shows the percentage of correct answers for each of the 12 LOs for the three semesters of our study. There appears to be a trend in correct answers for the three semesters. For example, in all three semesters, test questions for LO 8 (evolution) and LO 5 (respiration and photosynthesis) had the lowest percentage of correct answers, below 50% for all three semesters. This is not surprising, because these are three very difficult concepts for an introductory biology course.

Questions related to LO 7 (genetics) and LO 12 (communities and ecosystems) had the highest percentage of correct answers, above 60% for all three semesters; although LO 11 (adaptations to environment) was the only LO with correct answers above 80% for the last two semesters, its corresponding percentage of correct answers was slightly below 60% in fall 2010. At the introductory biology level, communities, ecosystems, and the environment tend to be easier concepts for the students to grasp, and this might explain the higher number of correct answers. Many of these concepts are introduced early in grade school (Ansberry & Morgan, 2007; Ballantyne, Fien, & Packer, 2001; Christenson, 2004; Santone, 2004; Smith-Sebasto & Semrau, 2004). The greater number of correct answers for LO 7 (genetics) was a bit surprising. Genetics is traditionally not an easy topic but is often one of the more interesting topics for introductory biology students. This may have led students to take the topic seriously, leading to a higher number of correct answers.

[FIGURE 1 OMITTED]

An alternative interpretation of the differences in correct answers among the LOs may be related to the design of the exam. Instead of LO 7 having more correct answers because of student interest in the topic, it could simply mean that the questions for this particular LO were easier than those for LO 8 or LO 5, which had fewer correct answers.

These results also led us to look at the LOs themselves. It wasn't until we were trying to design an exam with questions for each LO that we noticed that the LOs were so vague. We knew we needed questions pertaining to diffusion and osmosis, a common topic in every introductory biology course, but we had a hard time tying these concepts into an existing LO. We included these types of questions in LO 4 (Table 4), but by the very statement of this outcome, an instructor may be teaching only the types of cells and their structures and organelles. The awareness of the poorly written LOs led the faculty to begin a revision process. On approval by our College Curriculum Committee, the new LOs are incorporated in the spring 2014 syllabus.

Difference between postexam or final exam scores and grades

Figure 2 shows the differences between the final grades students earned and the grades they made on the posttests for fall 2010 and fall 2011, and on the common final in spring 2012.

For example, 38% of the students received a final grade of "A" in the course for fall 2010, whereas only 1% of the students earned an "A" on the posttest. Surprisingly, for the same semester, only 3% of the students earned an "F" in the course, even though 64% failed the posttest. These scores led us to examine the difficulty level of the exam. The numbers are very similar for fall 2011.

With the use of the common final in spring 2012 and the slight increase in the average score for the exam (61%), we expected to see a more positive correlation between course grade and exam grade. Even though the difference between course grades and exam grades was smaller in the spring 2012 semester, 38% of the students failed the exam, whereas only 16% of the students failed the course.

The following two considerations must be taken into account in interpreting these results. First, apprehension concerning grade inflation has always been a familiar quandary at most institutions, but the differences in a comprehensive final score and the course grades can many times be justified. Comprehensive finals encompass the entire semester, whereas final course grades are usually a combination of scores that concentrate on selected topics throughout the semester (Laurie, 2009). A grade of "C" or higher in BSC1005 (Introduction to Biological Sciences) is a prerequisite for BSC1010 (Biological Science I). Traditionally in this type of course, by the end of the semester many students who are earning a "D" or "F" have opted to withdraw from the course, resulting in fewer failing grades. In some instances, grades may include homework, projects, presentations, class participation, and/or attendance. Typically, students who maintain high averages throughout the semester do not need to make a high grade on the final to maintain their grade.

Second, the exam used in this study was created to include questions for all 12 LOs. Many of our colleagues admitted to not having time to cover all 12. Another consideration is the exam itself; the questions may have been too detailed, too difficult, or confusing in nature. While creating the exam, we also noted that the LOs were extremely vague and poorly written (Table 3), which may lead to varying interpretations as to which concepts and chapters should actually be included in the course. It is possible that not every instructor covers the same material during the semester; therefore, students in some sections may have been tested on concepts not covered.

Lessons learned

Our department's first attempt at creating a suitable assessment tool that could be used to improve our course was not without tribulations. What began as an attempt to evaluate student learning in our course quickly digressed into a complete alteration of our course LOs. Assessment must be an ongoing process; our results led us to begin the much-needed task of modifying our course. To recap, we generated the following list to aid us in any future attempts to create a common course final:

* Course LOs should be concise and clearly stated.

* Faculty should work together to create a detailed list of chapters and topics to be used as a guideline for instructors.

* Courses should not try to cover too much material.

* All faculty teaching the course should be involved in creating the exam.

* Creation of exam questions should be aligned with documented procedures.

* Protocol for deterring teaching to the exam should be seriously considered.

* Instructions for giving the exam should be clear.

* The importance of following instructions should be emphasized.

* Assessment is an ongoing process.

* LOs may need revisions.

* Exam questions may need revisions.

Conclusions

We found our study to be a valuable tool for helping us improve our course and the use of the common final a solid beginning in evaluating student performance. The results of our three-semester study are leading us to look at the course from a different perspective.

It is possible that the small increase from pre- to posttest scores may have been the result of factors other than the students' attitude toward the test. An alternative possibility for this small increase in grades may be the design of the course. With our current LOs (Table 3), almost every concept traditionally included in a general biology textbook is covered in a three-credit introductory course. The small increase in student performance may be because the amount of time was limited, each topic had to be presented in too general a detail to make sense to the students. It is possible that the instructors tried to hurry through the material to get as much information as possible in the short amount of time, rendering the course too difficult. In this case, in addition to the exam, a student and instructor survey may be warranted. The survey could specifically question the students' and instructors' perception of the amount of material introduced in the course along with the difficulty level.

The great difference between post or final exam scores and grades for the course led us to look at the difficulty level of the exam and also the possibility of grade inflation. We decided that, even though a common final can help evaluate student learning, it should not be used alone. Course grades are a combination of different assessment types and do not always correspond to the score of a single comprehensive exam.

By looking at the percentage of correct answers for each of our course LOs, we gained valuable information regarding the design of our course. The results of these data led us to question the number of topics that should be included in an introductory biology course and the detail with which we introduce these topics. It has also led to discussion and modification of the LOs, which were much too vague. Vague LOs make it almost impossible for instructors to teach at a consistent level. New instructors, especially, find it difficult to decide which topics to actually include in their lectures and how much detail is expected. By the beginning of the third semester of the present study, faculty were meeting to discuss revising the LOs, improving the exam questions, and offering workshops to share methods for teaching some of the more difficult topics identified in the study. Consequently, the study has served as a vehicle to stimulate faculty involvement in improving our introductory biology course. ?

Caption: FIGURE 1: Percentage of correct answers for learning outcomes (fall 2010 to spring 2012).

References

Ansberry, K., & Morgan, E. (2007). Into the woods. Science & Children, 44(8), 14-16.

Arum, R., & Roksa, J. (2011). Academically adrift: Limited learning on college campuses. Chicago, IL: University of Chicago Press.

Ballantyne, R., Fien, J., & Packer, J. (2001). Program effectiveness in facilitating intergenerational influence in environmental education: Lessons from the field. Journal of Environmental Education, 32(2), 8-15.

Christenson, M. A. (2004). Teaching multiple perspectives on environmental issues in elementary classrooms: A story of teacher inquiry. Journal of Environmental Education, 35(4), 3-16.

Elfrink, V. L., Kirkpatrick, B., Nininger, J., & Schubert, C. (2010). Using learning outcomes to inform teaching practices in human patient simulation. Nursing Education Perspective, 31(2): 97-100.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice guidelines for classroom assessment. Applied Measurement in Education, 15, 309-334.

Harris, D. N. (2011). Value-added measures and the future of educational accountability. Science, 333, 826-827.

June, A. W. (2011). Florida, with an eye on Texas, readies for next conflict over faculty productivity. Chronicle of Higher Education, 58(5), A17.

Laurie, R. (2009). Raising the bar: A data-driven discussion on grade inflation. Education Canada, 49(4), 32-36.

Liu, O. L. (2011). Value-added assessment in higher education: A comparison of two methods. Higher Education, 61(4), 445-461.

Mangan, K. (2010). Texas governor treats colleges like businesses. Chronicle of Higher Education, 57(6), A13-A16.

Mangan, K. (2011). Texas universities fear growing influence of governor's appointees. Chronicle of Higher Education, 57(33), A15.

Mohanty, G., Gretes, J., Flowers, C., Algozzine, B., & Spooner, F. (2005). Multi-method evaluation of instruction in engineering classes. Journal of Personnel Evaluation in Education, 18(2), 139-151.

O'Dell, I. (2009). Assessment and accountability in higher education. Educational Considerations, 37(1), 4-10.

Parrish, D. E., & Allen, R. (2011). An effective model for continuing education training in evidence-based practice. Research on Social Work Practice, 21, 77-87.

Popham, W. J. (2009). Diagnosing the diagnostic test. Educational Leadership, 66(6), 90-91.

Posner, M. A. (2011). The impact of proficiency-based assessment and reassessment of learning outcomes system on student achievement and attitudes. Statistics Education Research Journal, 10, 3-14.

Santone, S. (2004). Teaching for a healthier earth. Instructor, 113(6), 19-22.

Siebert, D. C., Siebert, C. F., & Spaulding-Givens, J. (2006). Teaching clinical social work skills primarily online: An evaluation. Journal of Social Work Education, 42(2), 325-336.

Smith-Sebasto, N. J., & Semrau, H. J. (2004). Evaluation of the environmental education program at the New Jersey School of Conservation. Journal of Environmental Education, 36, 3-18.

Snyder, C. A., & Carnicom, S. (2011). Assessment, accountability, and honors education. Journal of the National Collegiate Honors Council, 12, 111-127.

Tompkins, R., & Cates, L. (2009). Non-academic assessment in an accountability world: Achieving more with less. Community College Journal of Research and Practice, 33, 953-955.

Volkwein, J. F. (2010). The assessment context: Accreditation, accountability, and performance. New Directions for Institutional Research, 2010(Suppl. 1), 3-12.

Peggy Romeo (promeo@edison.edu) is a professor of biology and Frederick Posey is a professor of biology, both in the Science Department at Edison State College in Fort Myers, Florida.

TABLE 1 Differences between pre/posttest scores for fall 2010 and fall 2011. [DELTA] Year Exam Mean Median Mode SD N Mean 2010 Pre 45% 43% 43% 0.13 211 8% Post 53% 53% 47% 0.16 2011 Pre 40% 40% 40% 0.13 191 13% Post 52% 53% 50% 0.15 TABLE 2 Mean, median, mode, SD, and N for fall 2010 to spring 2012. Year Mean Median Mode SD N 2010 53% 53% 47% 0.16 211 2011 52% 53% 50% 0.15 191 2012 61% 63% 67% 0.19 387 TABLE 3 Introduction to Biological Sciences (BSC1005) learning outcomes (LOs). LO 1 Analyze basic atomic structure and function and discuss its role in chemical bonding. LO 2 Evaluate the roles pH, temperature, and enzyme catalyzed reactions contribute to metabolism. LO 3 Justify how the physical and chemical properties of water are important to life. LO 4 Compare the structure of prokaryotic and eukaryotic cells and discuss how structure relates to cell function. LO 5 Identify and analyze ways through which plant and animal cells obtain energy needed for metabolism and carbon needed for synthesis of molecules. LO 6 Analyze the stages and purpose of mitosis and meiosis. LO 7 Apply basic Mendelian genetics to solve genetic problems. LO 8 Analyze and evaluate the role of the evolutionary theory in uniting the various disciplines of biology. LO 9 Analyze and evaluate the applications and importance of advancements in genetic technology. LO 10 Explore the phylogenic relationships within major taxa of organisms. LO 11 Identify the relationships and adaptations of major taxa to their environment. LO 12 Evaluate the various types of relationships and processes within communities and ecosystems. FIGURE 2 Final course grades and final exam grades for the three semesters. Average Final Grades vs Average Final Exam Grades 2010 Actual Grade 2010 Test Grade 2010 A 38% 1% B 29% 6% C 20% 12% D 9% 17% E 3% 64% Average Final Grades vs Average Final Exam Grades 2011 Actual Grade 2010 Test Grade 2010 A 32% 2% B 26% 3% C 22% 14% D 14% 20% E 5% 60% Average Final Gardes vs Average Final Exam Grades 2012 Actual Grade 2010 Test Grade 2010 A 28% 4% B 27% 12% C 22% 17% D 7% 21% E 14% 45% Note: Table made from bar graph.

Romeo, Peggy^Posey, Frederick