Essays

Assessing Student Learning

Marshall D. Sundberg

Department of Biological Sciences, Emporia State University, Emporia, Kansas 66801

Biology education research has now reached a level of maturity where the expectation is that researchers will assess the effectiveness of their innovation on student learning. This may include an examination of affective outcomes, such as student attitudes and beliefs, as well as student understanding of discipline-based content. A variety of tools are available to generate assessment data, each with certain advantages and disadvantages. They include not only quantitative measures, which lend themselves to familiar statistical analyses, but also qualitative techniques that can provide a rich understanding of complex outcomes. This article describes some of the most commonly used assessment techniques, their advantages and disadvantages, and typical ways such information is reported.

Keywords: assessment, learning, biology education.

INTRODUCTION

Until recently, biology education articles focused on "how to" descriptions of classroom technique and laboratory exercises that the authors found to be successful in their own teaching. These might include applications of new technologies, both scientific and instructional, such as incorporating molecular techniques into laboratories and using a variety of computer-assisted instructional aids. Articles might also include descriptions of how new pedagogical approaches were implemented in individual courses, e.g., cooperative learning (Johnson *et al.*, 1991); concept mapping (Novak and Gowin, 1984); peer instruction (Mazur, 1997); investigative laboratories (Sundberg and Moncada, 1994); minute papers, (Angelo and Cross, 1993); and "Workshop Biology" (Udovic *et al.*, 2002) or "Studio Biology" (Montelone, personal communication). Success, however, was rarely defined beyond the instructor's impressions and student reactions. This has changed. We now recognize that student learning is often disconnected from the instructor's teaching. What is an elegant and rational presentation by the teacher may have little impact on the understanding of a majority of students. Worse, an instructor's presentation may have unintended outcomes that run counter to the intent of instruction (Sundberg, 1997; Sundberg and Moncada, 1994)! The only way an instructor can evaluate the effectiveness of an innovation on improving student learning is to plan and carry out a program of assessment.

The term "assessment" has a number of different meanings and connotations, but for the present purpose it may be defined as a systematic method to determine if, and to what extent, student learning has occurred. Unlike exams, whose purpose is to assign grades based on students' understanding, the purpose of assessment is to determine the impact of instruction on improving student learning. Although we are usually concerned with the end result of instruction, summative assessment, even more important are evaluations made during the course of instruction. Such formative assessment can guide changes as a course or program proceeds. Both formative and summative assessment can provide useful information about the efficacy of instruction, and today this kind of information is expected in reports and publications describing the activities (Stevens *et al.*, 1993; Frechtling and Westat, 1997).

One of the concerns that must be addressed in assessment is whether it is important to have a control, and what type of control would be appropriate. As scientists, we are accustomed to evidence that fits the natural science model of controlled experiments where procedures can be replicated and measurements reproduced. However, evaluation of the resulting data is interpretative (in fact, a qualitative assessment) and progress in science is the result of debate over these interpretations. Course size and intent of the study also raise some important questions. For instance, in large multisection courses it is relatively easy to match control and experimental groups, but to do so will require approvals by your campus Human Subjects Committee and individual student permission. With a small course an internal control may not be possible, although data from previous years (see "Database"

below) could be used for comparison. In fact, as the literature in biology education accumulates, other studies can provide a baseline for comparison (see for instance, Hake, 2002, for an example from physics education).

There are many different models of assessment techniques, but they can be divided into two basic categories, quantitative and qualitative. Each has important advantages and serious limitations. Therefore, it is important to consider assessment from the beginning of the study so that appropriate tools can be used, as seamlessly as possible, to generate meaningful data. Finally, it is now generally accepted that multiple assessment measures are required to adequately gauge student learning (Siebert and McIntosh, 2001). One source of general assessment information is FLAG (Field-tested Learning Assessment Guide) available online at http://www.wcer.wisc.edu/nise/cl1. Some of the most commonly used assessment techniques are described below.

QUANTITATIVE TECHNIQUES

Results of quantitative assessment are presented in graphs or tables in the same way as typical research data from a scientific study. A simple approach is to indicate mean change (Udovic *et al.*, 2002). However, there is less room for improvement if pretest scores are already high and there are frequently dramatic differences between sections. For instance, I once had two large sections of nonmajor biology students (more than 200 students per section), both of which had normally distributed pretest scores. However, the lower tail from one class barely overlapped the upper tail of the second! One of two techniques is usually employed to compensate for such initial differences. One alternative is to use an analysis of covariance (Sundberg, 1997). A second, simpler, approach is to compare "average normalized gain," observed improvement as a percent of the maximum possible gain (Hake, 2002). Some standardized quantitative instruments used for examining cognitive and affective attributes are listed by Tashiro and Rowland (1997).

PRETEST/POSTTEST

The most commonly used quantitative instrument is a content-based pretest/posttest. Typically, this is a multiple-choice instrument with questions written to address the major concepts of the course. The pretest, ideally, is administered prior to instruction, frequently during the first class meeting. The posttest is given at the end of the course and is usually the same instrument. Sometimes posttest questions are simply embedded into regular examinations. The questions are designed to fit the objectives of the course that may include more than simply content. For instance, the ability to analyze data and think critically are common objectives of general education courses, and these skills can be assessed using carefully constructed questions based on course content or independent of particular concepts that were covered in the course.

Advantages

Perhaps the main advantage of this type of instrument is that it is most similar to what we typically give as examinations, particularly in lower level courses. This provides both a level of comfort and a sense of reliability to the instructor. Because content is usually addressed, constructing these tests, and analyzing the results are similar to what we already do to assign student grades. The numerical scores can be analyzed statistically.

Disadvantages

That these assessments are similar to typical exams has a problem of "familiarity breeds contempt" both for the instructor and the student. The similarity to a typical exam makes it difficult to spend the time necessary to construct questions that adequately address the goals of the course. It also has the potential for student "abuse." How reliable are the data being generated? To a large degree, this depends on the character of the class and the rapport of the instructor. Pretests, especially in first-semester courses, typically have a higher reliability simply because of the naivete of the students. This looks like an exam and they tend to take it seriously. Posttest reliability can be questionable when students know that it will not affect their grade and a variety of attitudinal factors may influence their performance. This problem is sometimes addressed by embedding questions in a normal examination or by rewarding "gain" with "bonus points" that can raise a course grade. (The trade-off in the latter case is that students cannot remain anonymous, which therefore necessitates approval by the Human Subjects Committee if you intend to publish results.)

QUESTIONNAIRES

Quantitative questionnaires use a Likert scale, typically of 1–5, where students are asked to rank a statement based on a scale, for instance, from "strongly agree" to "strongly disagree." Again it is important to consider the goals of the course when developing the questions. Typically this kind of instrument is used to address goals that are not directly related to course content, such as student attitudes.

Advantages

The major advantage is the ease of obtaining information about student reactions and student perceptions as a result of the course. Another advantage is the ease of comparing quantitative scores from different sections or different courses.

Disadvantages

A well-designed instrument will be constructed so that for some questions a high score is desirable while for other questions the low score is preferred. The scores must be standardized before comparisons can be made. It is also important to use multiple questions to address each issue of concern to establish validity. As with the content-based posttest, one must also be concerned with how seriously students completed the questionnaire.

DATABASE

One reason to consider an assessment program at the same time that a course or curriculum innovation is planned is to provide a baseline against which student outcomes can be compared. Making significant changes to even a single

course can be very time consuming. A term or even a full year is sometimes required. During this time the traditional course may still be taught. By developing the assessment program as soon as the goals and objectives are set, it should be available to assess the effectiveness of the traditional course the last time or two it is offered. This establishes a baseline and the beginning of a dataset for future work.

Advantages

Once established, a database can grow to provide a longitudinal record of change. Ideally, a number of investigators at different institutions around the country could be adding to a common database that could provide real power to analysis of student learning. This has been done, for example, in the physics community where the Force Concept Inventory is a widely adopted assessment tool at two- and four-year institutions (Hake, 2002).

Disadvantages

The strength of uniformity has the disadvantage of not being tailored to a situation. Even in the same program, as a course or program evolves, it may be desirable to modify the instrument to reflect goals attained and new objectives that are defined.

QUALITATIVE TECHNIQUES

There is a general perception, especially among scientists, that qualitative assessment techniques are "softer" than quantitative tools—less accurate and less objective. In part this may be due to the inability to apply statistics to data analysis. However, good qualitative assessment is essential to understanding the complexity of student learning in the classroom. The best example of this is what we found in our early studies of using investigative-style laboratories to confront student misconceptions. Content-based posttesting suggested little gain in student understanding of natural selection (Moncada, 1993; Sundberg and Moncada, 1994; Sundberg, 1997). However, analysis of student interviews, journal writing, and concept mapping indicated that there were dramatic shifts in understanding of nearly half of the class. The problem was a nearly equal shift in the number of students who gained a more Darwinian view as there were who moved to a more Lamarckian understanding. We had no indication of this from the numerical data because the two trends balanced each other. The insight provided by the qualitative assessment drew our attention to unintended outcomes promoted by instructors that otherwise would have gone unnoticed.

The two major problems with qualitative techniques are that, first, they tend to be time consuming, and second, most of us have no training in their use. There are several solutions to these problems. Perhaps you have a colleague in the behavioral or social sciences, trained in using qualitative assessment, who would be interested in collaboration. If funds are available, a specialist could be hired, or in some cases commercial instruments with professional readers could be employed. In many cases, the best choice will be to use some of the simpler tools outlined below to gather limited amounts of very specific information.

There are several ways qualitative data can be reported. For example, representative examples of student work could be presented from early, middle, and later portions of a course (Moncada, 1993). Alternatively, representative answers to reflective questions can be used to illustrate common strengths and weaknesses (Udovic et al., 2002). Student responses can often be categorized by type and reported in the form of a table or graph (see Fig. 1 in Wright et al., 1998).

OBSERVATION

Observation is usually done by an outside assessor who observes and listens to student interactions during the course of instruction. The frequency of class visitation can vary, and the observer may choose to select and follow a specific cohort of students throughout a term or be more random in selecting subjects.

Advantages

There is little or no imposition on students or the instructor as data are collected during a class period. An added advantage is that instructor bias is removed from the assessment.

Disadvantages

This form of assessment is useful only in student-active situations such as laboratories or small classes where there is ample opportunity for individuals or student groups to interact with each other and with the instructor. It also requires that a qualified observer can be found, which in turn may be expensive.

INTERVIEWS

Interviews, like oral exams, provide an opportunity for you to probe a student's understanding of the material. It permits follow-up questions and interactions that can also provide insight into how a student is thinking and how thinking may change over time. In general, interview questions should proceed from more to less familiar material and from broad to more specific details. Especially in large classes, only certain individuals may be selected for interviewing. The individuals may be chosen at random, or specifically selected based on the objectives of the survey (Moncada, 1993).

One of the objections to involving the instructor in the interview process is the potential for introducing unintended bias into the assessment. The chemists at the University of Wisconsin (UW) devised a novel strategy to provide independent and unbiased assessment of student learning. Twenty-five faculty volunteers from math, engineering, and other science departments were asked to design their own 30-min oral examination over the chemistry course material. Each assessor interviewed approximately eight students (blindly assigned from two differently taught large lecture chemistry sections and octile ranking from the previous prerequisite course) and ranked the competence of the students they examined (Wright et al., 1998). Independently, researchers from the UW—Madison LEAD Center (Learning through Evaluation, Assessment, and Dissemination) used qualitative sociological research methods to assess the same students. There were significant two- and three-way correlations between rank in prerequisite class, grade in this class, and volunteer assessor relative ranking.

Advantages

The primary advantage of this approach is that it allows you to test thinking skills as opposed to content matter. It also allows you to follow changes in student thinking during the progress of the course (see introduction to qualitative assessment above). For the purpose of reporting results, a major advantage is that it can be used effectively in situations where there is no control group.

Disadvantages

The major disadvantage is the time commitment required from the interviewer. Of necessity, this will require scheduling times outside the normal class period.

FOCUS GROUPS

Focus groups of a sample of 5–7 students from a class provide an alternative to individual interviews. Because a certain amount of homogeneity is necessary to promote active participation within a group, three or four different groups are usually studied simultaneously to represent the range of diversity in the class. Focus groups are especially useful for uncovering attitudes, perceptions, and opinions. Questions should be planned in advance to be open ended and lead to specific objectives.

Advantages

Focus groups are good for identifying general patterns of student learning within a class. Open discussion between students can uncover information unanticipated by the instructor that can be very valuable in providing new insights.

Disadvantages

First, there is the logistical problem of finding a suitable time when all members of the focus group can meet. Second, a great deal of time and effort goes into planning a successful session and analyzing the results. Finally, this technique is not useful for uncovering details and specifics about individual learning.

CONCEPT MAPS

Concept mapping is a tool originally devised to investigate how students learn, but it can be used for a variety of different purposes including assessment (Mintzes *et al.*, 1999; Novak and Gowin, 1984). To construct a concept map, students must first identify the key concepts that were covered and then indicate the relationships between concepts. Concept maps have been used as a tool to establish a departure point during interviews.

Several different approaches have been used to evaluate concept maps. One approach is to establish knowledge categories ranging from common misconceptions about the subject to valid propositions. Initially, this may be based on a sample of 20 to 30 student maps that will become the baseline for a database (see above). Subsequently, individual student maps are matched to a category. Similarly, concept maps have been used to track developmental stage based on Perry's stages of intellectual development (Perry, 1970).

Advantages

One advantage of concept mapping over interviews is that maps can be generated simultaneously by an entire class. Concept maps also provide a permanent record of student understanding at a particular time, which is useful to show changes in student understanding.

Disadvantages

The main disadvantages of using concept maps are that, first, instructors must learn how to use and teach the technique, and second, students must be taught how to construct them—a process that can take up to a full class period. In-class use of the technique takes up instructional time.

JOURNAL WRITES, MINUTE PAPERS

There are several quick techniques, designed for formative assessment, that can provide useful information for reporting on the effectiveness of a teaching approach (Angelo and Cross, 1993). Among these are journal writes and minute papers. Journal writing is frequently assigned as an out-of-class activity where students are given a specific assignment based on that day's classroom activities. Journals are collected periodically and evaluated, similar to a laboratory notebook. Minute papers are an in-class activity, usually done at the end of class, where students are frequently asked to list or briefly write about the one or two most important concepts covered that day. This may be on a 3×5 card that is dropped off in a box on the way out of class. Minute papers can be read quickly (in extremely large classes, a sample of cards can be read) by an instructor.

Advantages

Two advantages are the low impact on class time and the relative immediacy of feedback to the instructor. The minute paper has the additional advantage of providing an attendance check in large classes.

Disadvantages

The information provided is narrowly restricted to the specific question asked and therefore focuses on only a small part of the material covered on any single day.

CONCLUSIONS

It is unfortunate that assessment currently is being mandated by accrediting agencies and the general public, because this reinforces the view of many scientist/educators that assessment is merely the latest educational fad to be forced upon us. Who can deny that the mandate is there, but in this case it is asking us to do in our teaching what we already do well in our science—test hypotheses about what will help our students learn. In my lab, if I am not satisfied with my results I make some modification and try again. This is assessment in action. We should apply the same skepticism about our students' learning to our classroom teaching as we apply to testing hypotheses in our research laboratory. This should be a natural progression.

Less natural for many of us is to accept that qualitative data can be as rich and informative as the statistical analyses we get from quantitative results. In fact, qualitative data are richer! Because quantitative assessment of necessity focuses on a few specific questions, but over a large number of students, it provides broad, generalized information about the class or program. Qualitative assessment, because of its open-ended nature, produces detailed information, but of a relatively few individuals. Quantitative assessment provides the broad strokes; qualitative assessment fills in the details. Both are needed to produce a good picture of student learning.

REFERENCES

Angelo, T.A., and Cross, K.P. (1993). Classroom Assessment Techniques: A Handbook for College Teachers, 2nd ed., San Francisco: Jossey-Bass.

Frechtling, J., and Westat, L.S., Eds. (1997). User-Friendly Handbook for Mixed Method Evaluations (NSF 97-153), Washington, DC: National Science Foundation.

Hake, R.R. (2002). Lessons from the physics education reform effort. Conserv. Ecol. *5(2)*, 28. Available on-line at: http://www.consecol.org/vol5/iss2/art28

Johnson, D.W., Johnson, R.T., and Smith, K.A. (1991). Cooperative learning: increasing college faculty instructional productivity, ERIC-ASHE Higher Education Report No. 4, Washington DC: George Washington University.

Mazur, E. (1997). Peer Instruction: A User's Manual, New York: Prentice-Hall. Available online at: http://galileo.harvard.edu

Mintzes, J.L., Wandersee, J.H., and Novak, J.D. (1999). Assessing Science Understanding: A Human Constructivist View, New York: Academic Press.

Moncada, G.J. (1993). Do college investigative laboratories really work? A descriptive study of an investigative biology laboratory in action. M.A. thesis, Louisiana State University, Baton Rouge.

Novak, J.D., and Gowin, D.B. (1984). Learning How to Learn. Cambridge, UK: Cambridge University Press.

Perry, W.G., Jr. (1970). Forms of Intellectual and Ethical Development in the College Years, New York: Holt, Rinehart, and Winston.

Siebert, E.D., and McIntosh, W.J. (2001). College Pathways to the Science Education Standards, Arlington, VA: NSTA Press.

Stevens, F., Lawrenz, F., and Sharp, L., (Frechtling, J., Ed.). (1993). User-Friendly Handbook for Project Evaluation: Science, Mathematics, Engineering and Technology Education (NSF 93-152), Washington, DC: National Science Foundation.

Sundberg, M.D. (1997). Assessing the effectiveness of an investigative laboratory to confront common misconceptions in life sciences. In McNeal, A.P., and D'Avanzo, C., eds., Student-Active Science: Models of Innovation in College Science Teaching, Orlando, FL: Harcourt Brace & Company.

Sundberg, M.D., and Moncada, G.J. (1994). Creating effective investigative laboratories for undergraduates. BioScience *44*, 698–704.

Tashiro, J., and Rowland, P.McD. (1997). What works: empirical approaches to restructuring courses in biology and environmental sciences. In McNeal, A.P., and D'Avanzo, C., eds., Student-Active Science: Models of Innovation in College Science Teaching, Orlando, FL: Harcourt Brace & Company.

Udovic, D., Morris, D., Dickman, A., Postlethwait, J., and Wetherwax, P. (2002). Workshop biology: demonstrating the effectiveness of active learning in an introductory biology course. BioScience *52*, 272–281.

Wright, J.C., Millar, S.B., Kosciuk, S.A., Penberthy, D.L., Williams, P.H., and Wampold, B.E. (1998). A novel strategy for assessing the effects of curriculum reform on student competence. J. Chem. Educ. *75*, 986–992.