Third Annual National Institute for Science Education
Indicators of Success in Postsecondary Science, Mathematics, Engineering and Technology Education: Shapes of the Future

Session One: Assessment in the Classroom

Assessment - a learning process: What evidence will we accept that students have learned?

Diane Ebert-May, Science and Mathematics Learning Center and Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona

As scientists, we spend a considerable portion of our professional time gathering information and making decisions based on that information. We read journals, attend conferences, and collaborate with our peers to obtain information that will improve our strategies for conducting research. We decide whether or not the strategies are appropriate for use in our research. As teachers, we also gather information, in this case, about our students. Based on that information, we make decisions about student learning and our teaching practice. We engage in the process of gathering and interpreting information and making decisions based on that information -- we assess.

The type of information we gather about our students depends on the evidence we will accept that the students have learned what we want them to learn. We must have confidence in the quality of the information if we want to justify our subsequent decisions about teaching. Major changes in assessment based on measurement theory and practice have catalyzed the development of new methods of data collection along with new ways of judging data quality. If, indeed, learning science should be an active process (Ebert-May et al 1997) then assessment should measure active knowledge (understanding, reasoning, and utilization) rather than discrete, isolated bits of inert knowledge. The new view is that "assessment and learning are two sides of the same coin" (NRC 1996, p 76). The methods used to collect educational data define in measurable terms what we should teach and what students should learn.

Hodson (1992) describes good assessment procedures as fulfilling at least four functions. First, a summative function: assessment should provide some description of students' levels of attainment in all components of the course. Second, a formative function: assessment should provide diagnostic feedback to the instructor and students throughout the course about the students' strengths and weaknesses, understandings and misconceptions in order to more effectively plan further learning by each student. Third, an evaluative function: assessment should provide instructors feedback about the effectiveness of the curriculum experiences provided in order to assist future curriculum decision making and planning. Finally, an educative function: assessment should engage students in interesting, challenging, and significant experiences aimed at helping them develop further insights and understanding. By using assessment as an educative tool, assessment becomes part of learning. Students actively participate in assessment and by doing so move towards taking responsibility for learning and advance themselves as independent, life-long learners (Angelo and Cross 1993).

The reform of assessment in this country is led by the K-12 section of the educational continuum (NRC 1996), and there is a critical need for higher education to join in this process. If we want to know how faculty can understand best what undergraduates are learning, retaining, and using in future contexts, we need to approach assessment with the same level of knowledge, rigor, and confidence as we do with the collection and interpretation of our scientific data. So too, approaches to implementing new forms of assessment should be comparable to utilizing new laboratory techniques. A scientist would seldom adopt a new laboratory technique without considering the purpose of the technique, the influence of the technique in the context of the experimental design, and the potential consequences of the technique in terms of results (Champagne and Ebert-May unpublished data). Change in assessment practices requires similar thought and examination.

Aligning assessment with learning: a case

Consider one of the student outcomes in our introductory biology course designed to develop biological literacy for all students. A goal was for students to effectively communicate an understanding of and links among biological principles and concepts to peers and others (Ebert-May et al. 1997). What evidence did we accept that students adequately communicated that understanding? To answer this question, we defined the tasks and performance standards appropriate for students in an introductory, non-major biology course guided by these questions: What type of written and oral communication assessment projects or tasks are appropriate? What are the biological principles that students must understand to communicate the ideas? What criteria would we use so both the students and instructors could differentiate levels of performance in both written and oral communication about the biological principles?

While our choice of assessment techniques to collect evidence about students' accomplishment of this goal was multifaceted, the underlying principle driving the choice of assessment was straightforward - the assessments we chose must enable students to communicate in both written and oral format their understandings of and links among biological principles. Simply stated, the choice of assessment form must be consistent with the student goal and what the instructor intends to infer from the data. The strategies we employed to address this goal included short writing samples, essay questions, quizzes, concept maps, self-evaluation, peer review of papers, class discussions, public hearings (Brewer and Ebert-May in press) and in the laboratory, oral presentations of research proposals, research papers, and poster displays.

For each of these assessment strategies we developed a scoring rubric. Scoring rubrics are the specification of the knowledge and ability components or the product characteristics as well as the point value assigned (Champagne unpublished data). Rubrics define the performance standards for a population of students based on a desired student outcome. For example, students addressed questions with short answers in class to communicate their understanding of and links among biological principles. For this task we defined the criteria for various levels of achievement (Table 1). Students were given this scoring rubric at the beginning of the course and were encouraged to use it as they wrote. We also solicited feedback from students regarding their understanding of and input to the criteria on this rubric. For other assessment strategies, students developed rubrics with their instructor. This increases further the learning that can result from engaging students in performance assessment.

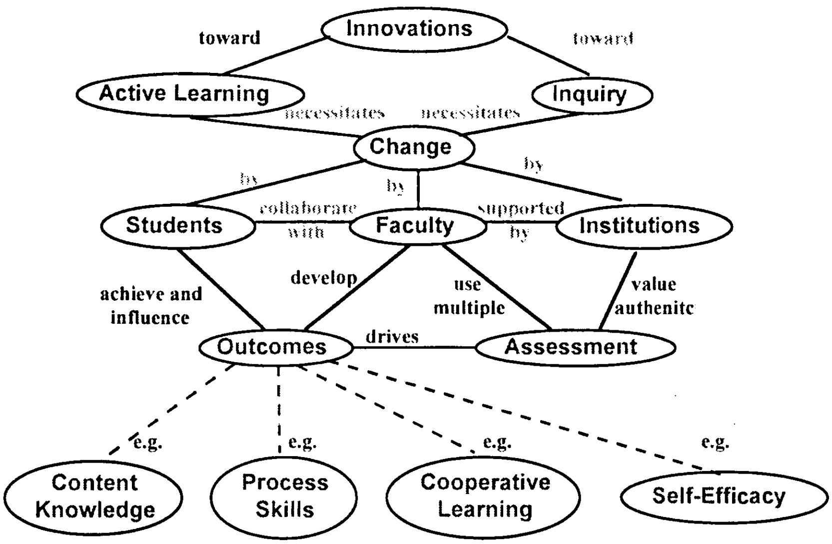
Table 1. Scoring rubric for short writing samples completed by students during class. Students were assessed on both their general approach and comprehension.

Level of Achievement	**General Approach**	**Comprehension**
Exemplary **(5 pts)**	• Addresses the question • States a relevant, justifiable answer • Presents arguments in a logical order • Uses acceptable style and grammar (no errors)	• Demonstrates a clear and complete understanding of the question • Backs conclusions with data and warrants • Uses 2 or more ideas, examples and/or arguments that support their answer
Adequate **(4 pts)**	• Does not address the question explicitly, although does so tangentially • States a relevant and justifiable answer • Presents arguments in a logical order • Uses acceptable style and grammar (one error)	• Demonstrates adequate understanding of question but does not back conclusions with warrants and data • Uses only one idea to support the answer. • Less thorough than above

Needs Improvement (3 pts)	• Does not address the question • States no relevant answers • Indicates misconceptions • Is not clearly or logically organized • Fails to use acceptable style and grammar (two or more errors)	• Does not demonstrate understanding of the question • Does not provide evidence to support their answer to the question
No Answer (0 pts)		

Our choice of assessment depended on the desired student outcome. For example, to gather formative feedback we used concept maps (Novak and Gowin 1984) and writing samples to assess students' understanding of the links among biological principles. So for a quiz, students were provided a list of concepts (5-6) and developed a concept map. Alternatively, as homework students were asked to identify the concepts for a set of readings and build a concept map. To prepare students for this type of assessment, we modeled the use of concept maps in class. Our intention was to use multiple forms of assessment to provide the kind of evidence we needed to make decisions about teaching. Figure 1 shows a concept map about innovation in teaching and learning that represents meaningful relationships between concepts in the form of propositions. The concepts are arranged hierarchically and provide a visual map to benefit both faculty and students.

Figure 1. Concept map representing the relationships between concepts about innovation in teaching. The concepts are arranged hierarchically, and the connection between concepts forms a logical proposition.



The nature of evidence faculty are willing to accept about student learning and the way they go about collecting and interpreting the evidence provide the best guide to inform their decisions about teaching. Importantly, the student goals and assessment must be based on acceptable, well defined criteria that faculty will accept as evidence of student achievement. If faculty value students' abilities to solve interdisciplinary problems in biology, assessments must include opportunities for students to demonstrate their ability to solve problems. This is not to minimize the importance of understanding content. Rather, the issue is how to assess students' knowledge in ways that demonstrate their in-depth, long-term understanding of content.

Self-evaluation: a way to the reduce risks of innovation

Problem

During this session, several interrelated themes focus on the nature of evidence we gather and accept about student learning, and how our interpretation of that evidence influences what we do in a classroom. Assessment often leads to innovation in the classroom, but innovations can create disequilibrium between student expectations and what occurs in a classroom. Many students still expect to be passive learners in an instructor-centered classroom where faculty talk and students listen (Ebert-May et al 1997). Furthermore, new undergraduate students view knowledge as existing absolutely and concretely, and assume knowledge to be certain (King and Kitchner 1994). Students entering college are dualistic thinkers (Perry 1970). They are intolerant of ambiguity, answers are right or wrong, black or white, they either get an idea or do not get an idea. Students are not prepared to try to understand (Belenky et al 1986). Therefore, it is not uncommon that faculty who are innovative and who attempt to create active learning in the classroom show a measurable decrease in teaching evaluations if the types of evaluations used have little to do with the reformed goals of the course. For example, student evaluations required for each class in the College of Arts and Sciences at my university include the following questions (Table 2).

Table 2. Student Evaluation: Course and Instructor Characteristics. For each item, students bubble in their choice on a scale of 1 (strongly disagree) to 5 (strongly agree):

1. The objectives of the course were made clear to me.
2. The instructor accomplished course objectives.
3. The instructor seemed genuinely concerned with student progress.
4. My interest in the subject has been stimulated by this instructor.
5. The course was intellectually challenging.
6. My general estimate of this course.
7. General estimate of this instructor.

The mean point values for these questions (based on a scale of 1-5) are tallied, and individual faculty scores are compared to the college mean, usually 4.0 with a standard deviation of 1.0. Although faculty are encouraged to design and use other types of course evaluations, the reality is that the mean point score from this instrument becomes an important variable used during discussions by promotion and tenure committees about a person's teaching. Furthermore, cases have been reported in which a faculty member's mean score was 3.9 compared to the College mean of 4.0 and, therefore, this individual's teaching was considered below average. Faculty ignored interpretation of the mean and standard deviation in this case.

How does the student assessment instrument in Table 2 provide substantive feedback to faculty about the goodness of fit between student learning goals, pedagogical methods, and student learning styles? It does not. Furthermore, non-instructive negative feedback to faculty fails to encourage or reward thoughtful innovation and careful analysis about teaching and learning. It takes less time and involves fewer risks to teach in a traditional manner (with enthusiasm) and get higher point scores. We all know how to do that. Freire describes traditional education as "banking" -- the instructor's role is "to 'fill' the students by making deposits of information which the teacher considers to constitute true knowledge" (Freire 1971, p.63). The students' job is merely to "store the deposits."

An alternative solution

Non-instructive student course evaluations should be reconsidered. Students need to move towards a mode of reflecting and evaluating their own understandings and abilities and providing that feedback to faculty. To do this, we began utilizing student self-evaluations about course goals as a measure of learning and as substantive feedback to guide changes in faculty practice.

Self-evaluation is integrative, reflective work throughout a course that emerges as an ongoing process through various assessment strategies (Angelo and Cross 1993). Through this process, students reflect on and evaluate their own scientific understanding and ability. When students reflect on their accomplishments in a course, they are really conducting a self-evaluation. It is both a process - students think and write about what they did and learned in a course, and a product - it is a written document that informs faculty about how students regard their accomplishments and how they achieved those accomplishments.

Student self-evaluation is both an old approach and a new one (MacGregor 1993). The method was used historically in alternative colleges, and eventually more traditional colleges and universities began to use self-evaluation to engender student active participation in the process of evaluating their learning. Learning theorists strongly advocate the value of having students think more reflectively about what they know and what they can do. (MacGregor 1993). As students gain experience in this, self-evaluation becomes an important learning strategy as well as an avenue for alternative forms of assessment. While enriching learning for students, student self-evaluations also can help faculty and departments learn about student learning (MacGregor 1993).

Feedback from student self-evaluations informs us about what keeps students motivated, engaged, and interested, and about what they consider important in the course (MacGregor 1993). What we read may provide us insight into the teaching and learning going on in our classrooms. Perhaps what we thought students learned, they did not learn at all; what we thought was clear and simple, was complex and confusing to students; what we predicted to be a successful pedagogical strategy, was perceived as cumbersome by students (Kusnic and Finley 1993).

Our research indicates one of the most important effects of self-evaluation is that it is a strategy that actively engages students in their learning - in terms of not only what they can do, but also in terms of what they cannot do at this point in time, what directions their learning must take, what must they do better.

"Students who internalize valued achievement targets so thoroughly as to be able to confidently and completely evaluate their own and each other's work, almost automatically become better performers in their own right." (McMillan and Forsyth 1991)

Self-evaluation: the process

We used self-evaluation in a large introductory biology course for non-majors with an enrollment of over 600 students. The self-evaluation was designed specifically for the students to comment on their accomplishment of the goals of the course. It was a written assessment in which the students provided both quantitative and qualitative synthesis of their learning. Students put themselves in the center of the learning experience, rather than focusing on the instructor. Writing self-evaluations was a challenge for students and required guidance, practice and time.

We included the assignment and rationale for the self-evaluation in our syllabus. Students were informed on the first day of class that they would gain a perspective about where they have been as a learner and knower throughout the course, and would need to consider what they needed to do next. To accomplish this, we integrated formative reflective work throughout the course. For example, periodically we asked students to write short statements about the function of their cooperative groups. Alternatively, we would ask them to reflect on the effectiveness of an assessment strategy, such as concept maps, on their understanding. Importantly, we provided some type of feedback to student responses.

As a summative assessment, we asked students to provide us a self-evaluation at the end of the course that described their perception of their accomplishments in the course as well as the accomplishments of the faculty. The instrument was based on student and faculty goals for the course, and each question had two parts. First, for each outcome students were asked to indicate on a five-point Likert scale the degree to which they accomplished each goal. Then in the space below each outcome, students were required to explain what happened during the course that influenced the choice they circled.

The instrument was on the www, and students had ten days to complete the assignment. Students electronically submitted the assignment, which included their names, and they were assured that we would not read the comments until final grades were assigned. Students received points for completing the self-evaluation, either all of the points or none. Maximum points were awarded to each student who completed the instrument fully, that is, written comments that included statements, examples and backing. No points were awarded if any information was missing or if a student merely repeated the question. A graduate student quickly examined each paper for completeness and awarded full credit or no credit. No students voiced concerns about anonymity since an environment of trust had been developed during the course. Furthermore, since the nature of the guiding questions for self-evaluations focused on the learner, students tended not to "blame" successes or failures on the instructor. The majority of comments focused on student perceived accomplishments that often included likes or dislikes about the course in the context of their achievement.

Table 3. Self-evaluation instrument for students in introductory biology. Each question includes a five-point Likert scale and an explanation section for extended responses as shown in Question 1.

1. To what degree have you increased your ability to describe how other people have used the process of science?

High Degree Moderate Degree Somewhat Minimal Degree Not at all
Explanation:

2. To what degree can you effectively communicate an understanding of and links among biological principles and concepts to peers and others?
3. To what degree have you developed confidence in your ability to write about, criticize and analyze concepts in biology?
4. To what degree have you increased your ability to use the process of scientific inquiry to think creatively and formulate questions about real-world problems?
5. To what degree have you developed positive attitudes about the relevance of biology to your life and the ability to apply this knowledge in the resolution of real-world problems?
6. To what degree have you enhanced your understanding of biological concepts and application of them to personal, public, and ethical issues?
7. To what degree have you enhanced your ability to reason logically and critically to evaluate information (i.e., be skeptical)?
8. To what degree did you develop positive interdependence and individual accountability within your cooperative groups?
9. To what degree did the instructor provide a learning environment in which all students participated in a variety of instructional strategies and assessment practices that challenged your higher order thinking and reasoning skills so you could successfully demonstrate the outcomes described above.

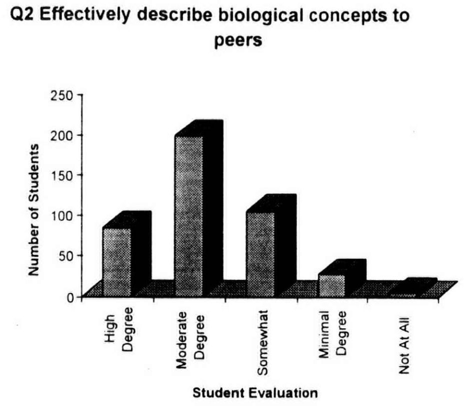
Frequency of each response was tallied for each item. Then the written comments associated with each item were coded and interpreted using NUDIST software (Non-Numerical Constructed Data, Indexing Search and Theory Building, QSR: Qualitative Solutions and Research, Inc. 1997). The combination of quantitative and qualitative responses provided two complementary approaches to interpreting trends in the data, a technique commonly used in science education research (Lancy 1993).

Self-evaluation: results and interpretation

Figures 2-5 show the frequency of responses students provided regarding the degree to which they thought they achieved four of the goals for the course. The most frequent student response was that they accomplished the goal to a "moderate degree," with over 90% of the students reporting they achieved the goal to a high degree, moderate degree or somewhat. We did not expect the majority of students to respond in the "high degree" category because the course was challenging to most students. Rather, we predicted that most students would recognize that they needed to know and do more to accomplish each goal to a high degree. The explanations students wrote after each choice confirmed this prediction. Students recognized they were making progress towards accomplishing the goal yet, at the same time, realized that they needed more experience, practice and knowledge to become highly effective in achieving the goal. Students were realistic about what they could accomplish in a one-semester course, but also described what they needed to continue to learn in the future.

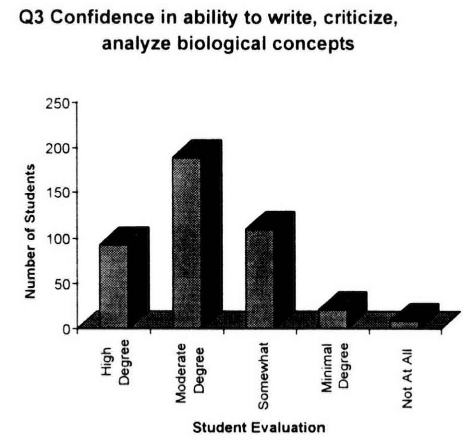
The sample quotations below Figures 2-5 represent the types of self-reflections commonly provided by students. We used these statements to help interpret the frequency distributions for each question, a combination of quantitative and qualitative assessment. For example, although all cooperative learning groups were not perfect (Figures 2 and 5), the majority of students wrote about the value of group work to their learning and provided specific examples of how their group functioned to achieve their goals. Alternatively, reasons for less effective cooperative groups were provided with specific examples. The comment in Figure 3 suggests that the extensive writing done in class was useful to students, although not necessarily their favorite task, "...We were forced to write down how..." Various performance assessments showed that students' writing improved significantly throughout the course. For example, students explained why the position papers and associated public hearings were a meaningful assessment strategy (Figure 4), therefore we continued to use position papers and public hearings as an assessment strategy in the course.

Figure 2. Frequency of responses and sample explanation to the goal: To what degree can you effectively communicate an understanding of and links among biological principles and concepts to peers and others?



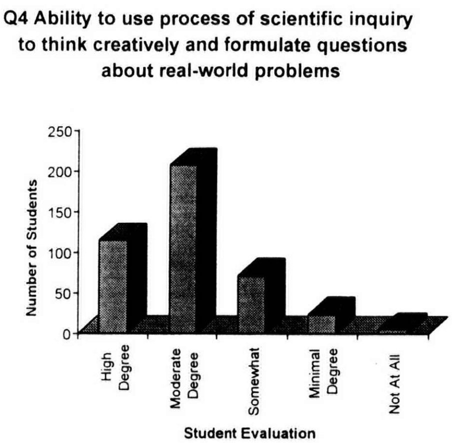
"So much group work made me realize that understanding a concept and being able to communicate a concept are different things."

Figure 3. Frequency of responses and sample explanation to the goal: To what degree have you developed confidence in your ability to write about, criticize and analyze concepts in biology?



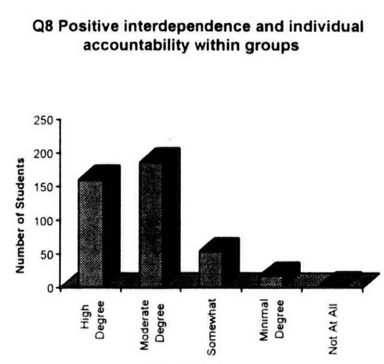
"I always was good at biology (or at least got good grades in biology), but never really understood any CONNECTIONS. This class made it clear to me that everything is connected to each other. We were forced to write down how we UNDERSTOOD concepts, not simply to memorize parts and functions."

Figure 4. Frequency of responses and sample explanation to the goal: To what degree have you increased your ability to use the process of scientific inquiry to think creatively and formulate questions about real-world problems?



"...we did our position papers on two real world issues that we deal with every day. One paper we did was on the environment and the other was on breast cancer...This class really makes you think about how important scientists are in the world today. They try to formulated and answer questions that will help us survive in the future..."

Figure 5. Frequency of responses to the goal: To what degree have you enhanced your ability to reason logically and critically to evaluate information (i.e., be skeptical)?



"Working in groups requires so much responsibility. This means coming to class every day, having your input for all of the questions and quizzes and most of all showing up on your own time to complete homework assignments, etc.... Working as a team requires dedication and cooperation This is how everyone will succeed."

In our course, we used student self-evaluation in the same way we required students to reflect about concepts when they compared and contrasted ideas, analyzed their work, or explored the implications of a theory. Both required reflective thinking. During the process of self-evaluation, students' learning moved from a passive process to an active, meaningful process (Kusnic and Finley 1993). King and Kitchner (1993) consider evaluation one of the higher level thinking skills, along with analysis and synthesis, that comprise "critical thinking." Hence, we incorporated self-evaluation into our course as one way to begin to move students along the continuum of intellectual development from concrete thinking to higher-level thinking. We considered self-evaluation a learning strategy that helped students construct meaning of concepts, derive relevance of ideas, and begin to build a coherent framework for continued learning.

Assessment is a learning process. Both faculty and students benefit from meaningful assessment information. Self-evaluation is one strategy to reduce the risk of testing and refining innovative teaching practices because it provides faculty direct information from the students about their perceived achievement of goals. Importantly, we used student self-evaluation as an alternative to poorly designed student evaluations that neither informed nor modeled best teaching practice.

References Cited

Angelo T, Cross KP. 1993. Classroom assessment techniques: a handbook for college teachers. San Francisco (CA): Jossey-Bass Publishers.

Belenky MF, Clinchy BM, Goldberger NR, Tarule JM. 1986. Women's ways of knowing. New York (NY): Harper Collins Publishers.

Brewer C, Ebert-May D. in press. Dynamic discussion in a large lecture hall: the case of genetic engineering. Journal of College Science Teaching.

Ebert-May D, Brewer C, Allred S. 1997. Innovation in large lectures-teaching for active learning. Bioscience 47: 601-607.

Freire P. 1971. Pedagogy of the oppressed. New York (NY): Seaview Press.

Hodson, D. 1992. Assessment of practical work. Science and Education 1: 115-144.

King PM, Kitchener KS. 1994. Developing reflective judgment: understanding and promoting intellectual growth and critical thinking in adolescents and adults. San Francisco (CA): Jossey-Bass Publishers.

Kusnic E, Finley ML. 1993. Student self-evaluation: an introduction and rationale. in MacGregor, J., Ed. Student self-evaluation: fostering reflective learning. San Francisco (CA): Jossey-Bass Publishers.

Lancy D. 1993. Qualitative research in education: an introduction to the major traditions. White Plains (NY): Longman Publishing Group.

MacGregor J. 1993. Student self-evaluation: fostering reflective learning. San Francisco (CA): Jossey-Bass Publishers.

McMillan J, Forsyth DR. 1991. What theories of motivation say about why learners learn. in Menges RJ and Svinicki D eds. College teaching: from theory to practice. San Francisco (CA): Jossey-Bass Publishers.

National Research Council. 1996. National science education standards. Washington (DC): National Academy Press.

Novak JD, Gowin DB. 1984. Leaning how to learn. New York: Cambridge University Press.

Perry WG Jr. 1970. Forms of intellectual and ethical development in the college years: a scheme. Troy (MO): Holt, Rinehart & Winston Publishers.