Testing Alternative Assessment Strategies—The Ups and the Downs for Science-Teaching Faculty

Assessing Assessment—Some Surprises and Recommendations from Florida State University's Assessment-Evaluation Study

Hedy Moscovici and Penny J. Gilmer

Under part of a National Science Foundation (NSF) grant designed to restructure science and science education courses, and in the atmosphere of the general reform movement in education, faculty at Florida State University (FSU) set out to develop a new biology course for prospective elementary and early childhood teachers. We set as our goal a restructuring of the science-content courses to help prospective teachers of these courses become confident and comfortable with scientific concepts and processes and be able to use these concepts in their classrooms and lives.

The development of the biology course involved interdepartmental efforts. Faculty and graduate students from arts and sciences and from the science education department worked with district representatives and elementary teachers in the course development team meetings. We discussed questions and main issues during these meetings, which were similar to those described by Caprio et al. (1989), namely, motivation, selection of the main concepts to teach, how to relate to the different models and abstractions, how to make the topics more relevant, how power relationships in class will influence students' learning.

Hedy Moscovici is assistant professor, department of biology, Western Washington University, Bellingham, WA 98225. Penny J. Gilmer is associate professor, department of chemistry, Florida State University, Tallahassee, FL 32312.

Twenty student; participated in the USF assessment study. Some spoke of the "busyness" of the program, citing the numerous assignments.
how to develop the material in an inquiring mode, and how to overcome students’ past negative experiences with science.

There were two differences between our study and that of Caprio et al.: first, we had a more restricted student population, namely prospective elementary and early childhood teachers and, second, we allocated a relatively large portion of the development time to discussion of alternative modes of assessment. While individuals with a strong science education background supported the idea of using diverse modes of assessment, biology faculty resisted most alternatives, opting instead for traditional tests and quizzes with some alterations.

This paper addresses the different alternative assessment strategies used by the six different biology instructors, each teaching a section of the special class of 22 prospective teachers. Our paper also provides suggestions for similar relatively small classes, as well as large college classes.

ALTERNATIVE ASSESSMENT

There is a vast literature exploring the meaning and purpose of assessment. Pollio and Humphreys (1988) recognize the fact that grades represent “judgments made by human beings about complex processes” (p. 95), and that they should be used only in the class context. Rutherford and Ahlgren in Science for All Americans (1990) emphasize the need to modify assessment instruments “so that they could become incentives for purposeful learning” (p. 212). Such a change calls for individuals to “invest more heavily than in the past in developing new kinds of tests to provide practical alternatives to tests that reward only the memorization of bits of information” (p. 212).

Fulfilling the Promise: Biology Education in the Nation’s Schools (National Research Council, 1990) broke with the tradition of standardized tests, and stated that the “understanding of central concepts and principles in biology will not be gained as long as classrooms and standardized tests assess only recall and recognition” (p. 45). In the recommendations, the authors ask for individuals to publish “model examinations” that can measure “cognitive and affective changes and theoretical and applied knowledge” (p. 46) and from publishers of tests that accompany textbooks “to improve the diversity and quality of their tests” (p. 47).

Collins (1993) provides a clear definition of alternative assessment, and underlines the advantage of using a multidimensional tool for assessment:

Alternative assessment is not a single entity; rather it is a collection of modes of gathering data to describe what students know and are able to do. While these modes of data gathering share different characteristics, no one assessment possesses all the characteristics (Collins, p. 123, 1993).

The idea behind alternative assessment strategies lies with the recognition of the diversity of the student population and the concomitant ways to express learning. Assessment involves subjectivity on the part of the instructor in deciding what and how to assess and grade the students. Working with the philosophy of alternative assessment means making a shift from the instructor’s subjectivity and control over the questions on the test, the allocation of points, and the translation of percentages into letter grades. It means taking the direction toward a system instead that acknowledges students’ subjectivity in terms of diverse ways to express learning.

In our case, each of the alternative modes of assessment (journals, portfolios, laboratory projects, performance-based tests, etc.) was introduced mainly by the science education group to the biology faculty during the development team meetings. These alternatives were described in terms of entries that could provide opportunities to identify students’ learning from different angles. Their combinations provided the necessary variety for all students to be able to demonstrate their own learning.

With a restricted student population (22 students), the instructors in this course had the opportunity to experiment with a variety of assessment techniques. The syllabus contained some of the suggested alternatives (such as using portfolios to provide selected evidence of learning and student participation) with a negotiated percentage of the final grade attached to each (i.e., 15 percent for portfolios and 10 percent for class participation).

As a result of discussions during the development stages, we utilized additional alternative assessment strategies as well as the traditional and modified tests and quizzes. These alternatives included the use of journals in which students could address unclear topics, formulate their understandings in terms of concepts developed during class, and suggest ways to improve the course. Other alternatives took the form of hands-on activities, sheets from laboratory activities and from field trips, in-class presentations, and performance assessment, such as using a microscope to identify specific parts of an organism. The faculty tried the multiple-choice test in a “take-home” version, in which students answered examination questions by themselves but in which they could use books and notes. Several instructors used essay questions as a variation to the traditional multiple-choice test.

DESCRIPTION OF THE STUDY

Twenty-four prospective elementary and early-childhood educators registered for this experimental course in biology. The course promised in its advertising to have the advantage of having a small class, to take into account the special class population, and to enhance learning by using a variety of hands-on activities.
Two students dropped the class, which continued with an enrollment of 22. Twenty of the students were females, and two were males (a usual profile for the population of students having this interest). All but two of the students were freshmen, and those two were sophomores.

Six instructors were responsible for different sections of the same course, and a laboratory assistant took care of all the laboratory preparations and the smooth development of the different parts of the course.

Class met three times a week during one semester for nearly two hours per session. A Saturday educational trip to a marine laboratory in the area was added to the general schedule. The course counted for four credits of science requirements.

The different instructors decided on the assessment strategies to be used in their individual section. When two instructors had to teach the same section, each one provided a grade, and they decided as a team on the distribution of percentages according to the amount of time each instructor had with the class.

All class meetings were videotaped, and transcribed partially or as a whole. Interviews with instructors and students in class were also transcribed and used in the process of the class analysis. Different artifacts (syllabi, exercise papers, quizzes, tests, overheads, etc.) also added in the process.

RESULTS

In presenting the results and in discussions on the use of alternative assessment strategies, we have focused on two main difficulties that led to discomfort, and a poor fit between students' expectations and their actual experience in class. One student who dropped the course during the first week did so because of the large amount of content, assignments, and experiments expected from the students. Other students also expressed their discontent with "lots of work" that they had to do, and the "lot of effort" they had to invest in learning for the examinations because of the large amounts of information presented. Collectively, students underlined the "busyness" of the course—there was always an assignment due in the form of a quiz, an examination, a filled-in diagram, a part on an ongoing project for class or for the laboratory parts, journal time, or a lab report. The multiple use of alternative assessment strategies was not fulfilled in terms of grade representation for students' learning because of instructors' unfamiliarity with theoretical aspects of these strategies, and their need to transform "subjective" entries like students' participation or portfolio assessment into "objective", quantifiable entries. These two very important issues are discussed below:

▲ Instructors' unfamiliarity with alternative assessment strategies. Instructors' unfamiliarity with the alternative modes of assessment suggested a lack of commitment to implement what was proposed and accepted during the development team meetings. There were seminars and paper-sharing sessions before and during the implementation of the course. The sessions were designed to provide opportunities for the instructors involved in the restructured course to build the necessary background vital to the understanding of the meaning of alternative assessment strategies. In most cases, faculty from arts and sciences resisted learning new techniques for teaching and assessing biology. Discomfort with alternative assessments was sometimes verbalized. An example of this—looking at the portfolios at the end of the semester and deciding to award the entire 15 percent to all the students who submitted a portfolio because some faculty found it difficult to try to develop criteria to assess the portfolios presented by the students. Other instructors ex-
pressed discomfort in a more tacit way by not even trying to evaluate portfolios and reverting to evaluating the students solely by tests and quizzes.

Transformation of "subjective" assessment into "objective" assessment. Some of the faculty expressed the need to transform what they called "subjective" assessment into an "objective" form in which they established the criteria. Lack of understanding of the meaning of alternative assessment strategies forced some faculty to transform the different students’ tasks into "objective," comparable, and easily quantifiable ones. This was done without regard for students' perceptions of their learning; and, in most of the cases, led to discontent and frustration in the students.

Many students felt that their grades did not reflect their efforts and commitment and, more significantly, did not represent their learning of biology. As an illustration of this assertion, we provide the discussion between one student who received a C in one section of the course, and one of the two instructors who was in charge of that section.

Tara [crying]: It’s not worth it! I could sit in the regular BIO 1600 and make an “A”.
Dr. Landers: [showing Tara her grades in her notebook]: You see? It is the average. You have two As on first two quizzes, while each counts for 4 points, 90 percent on the take-home test that was worth 100 points, 3 from the 5 points on the last quiz, and an S [satisfactory] on the group project. The problem is that you have got only 72 percent from the 100 points allocated on the final exam. You see? It makes a total of 173 points and that’s a C.
Tara: But I studied so hard together with student Y. For three days till late!
Dr. Landers: I understand, but what can I do? You see, there are students who got “A’s”, and “B’s”, as well as Cs, and even Ds.

Dr. Landers decided on the number of points for each of the assessment activities, and although she used many alternative assessment strategies, the tests counted for 200 of 213 points (94 percent).

One of the assessment tasks was a take-home test, the other an in-class one, and the questions varied from the multiple choice to the fill-in-the-blanks, to a few essay type, asking for short answers. Although the grading system was Dr. Landers’ creation, when faced with the results she seemed powerless, saying “... what can I do?”

She reassured herself when students received grades between “A” and “D” in a way that approximated a normal curve. Dr. Landers constructed a relationship between percentage of points and the grade without taking into consideration students’ criteria regarding their own learning. The university bulletins did not provide any explicit correlations between percentage of the achievement and the letter grade, so individual faculty had the freedom to make their own decisions.

It is interesting to observe that many of the students regarded the hands-on activities as valuable learning experiences. For example, the students regarded building the “Biology Belle” (i.e., a stand on which different groups of students built the different body systems into a whole, integrated human body) as a very good, integrative exercise. Pam, one of the students in class, expressed her valuable learning experience using the “Biology Belle” project as following:

I knew the names of the various body systems and the elements that composed them, but I didn’t understand how they fit together within the human body. This project allowed hands-on creative experience and the opportunity to put the body systems together and form a person. We made the esophagus from a long straw that went into the leather pouch (the stomach), and a pantyhose stuffed with yarn filled the place of the intestine. We had to position the body parts above the pelvis bone, but below the chest to make room for the respiratory and circulatory systems (Moscovici, 1994, p. 72-73).

The students only received a grade of S/U for this activity, with all students receiving an ‘S’, but the assessment did not translate into points toward the final grade. Because of the multitude of assignments and tasks, students did not have enough time to present their part to the whole class, and the project lost its educational potential. After the course, we discussed with Dr. Landers the learning potential of the “Biology Belle” project, and she decided to allow more time for students to share their part with the class in future course implementations. She also expressed the need to re-address content topics and number of assignments as their amount was relatively high, not allowing enough thinking and application time.

The decision to have grades for each of the four units of the course forced instructors from each team to work together to decide on common values for the assessment process. Individual faculty did not discuss assessment strategies with each other prior to this class. That is why they were surprised to find that some instructors assessed students’ presentations on an S/U basis, while others used student peer assessment in the process of determining a grade. Some instructors gave credit to students who linked personal experiences to biological concepts, while other instructors only gave credit when students’ responses matched the expected responses.

Looking at the assessment criteria, written assignments counted for more points than oral presentations, even though these students had already decided on a teaching career that will rely on their ability to communicate orally what they know about biology. The only category that could assess students’ ability and interest as expressed orally was “participation.” Unable to use such a subjective entity, the instructional team used “attendance” instead because attendance was easier to quantify.
With this transformation, the faculty did not evaluate the ability of the students to speak of their understanding of biology. There were many cases of students who attended but did not take an active part in the lesson who received more points on “participation” than other students who, although they attended fewer sessions, were actively involved in answering and asking questions, or providing examples.

Although considered an “objective” entity, the grade on each unit reflected the values of the instructors rather than what the students learned in terms of scientific concepts that they could utilize in the future.

Instructors with some background in education or science education were more open to new ideas and had the will to re-think assignments. For example, Dr. Landers decided to provide students with diagrams of the body systems to be used as resources instead of items to be returned as assignments. Instructors without a background in education or science education did not show interest in the alternative assessment strategies, and did not attend meetings that were geared toward their clarifications and explanations.

SUGGESTIONS FOR OTHER SIMILAR COURSES

Grading has to reflect not merely assignments that are easy to evaluate but students’ learning. As instructors, we have the responsibility to be receptive to feedback from the students and to know when and what they learn, and how they may use what they have learned in the future. It may involve, for example, observing students during laboratory sessions, evaluations of oral presentations (by peers, self, and instructor), or/and portfolios that provide evidence of a student’s learning.

Instructors need to become familiar with alternative modes of assessment by learning and reflecting on alternatives prior to the implementation of the course. They should be aware of the educational principles in alternative assessment strategies in which recognition of individual diversity replaces the instructor’s subjectivity.

In our case, it would have been very helpful to share information regarding alternative modes of assessment, understand them and their rationale, and decide as a group, or on individual units what to use, and to what extent.

Instructors must commit themselves to the use of assessment once they decide to use them, even if the assessments are more time than expected. In the case of alternative assessment, even though it may take more time, it provides a more accurate image of student learning and the students’ ability to use their understandings in the future. The question of appropriate assessment should address the extent to which the assess what assessment students know rather than efficiency in terms of time spent per student.

SUGGESTIONS FOR USING ALTERNATIVE ASSESSMENT STRATEGIES IN LARGE LECTURE SETTINGS

- Use journal entries so students can express their understandings, as well as any unclear issues remaining after class. As it is difficult to address all the journal entries at one time, one possibility is to select a manageable number (i.e., all the social security numbers ending with the number X) and rotate.
- Use minute papers as described by Zahn (1991), in which he wants to know if students’ have their expectations met in class, and when they are not met, whose “fault” was it?-instructor’s or student’s? He wants to know the topics that remain “the mudiest.” It is implicit that the teaching and learning in class are the dual responsibilities of the instructor and the students.
- Use electronic mail messages to converse with students. Again, the density of messages can be controlled by the instructor depending on the amount of time and help available (Brabston & Krishnan, 1993).

- Have the students work in groups on applications of different concepts than were presented in class. The different projects could be examined and the best, let’s say five projects would get 2, 4, 6, 8, and 10% bonus points added to students’ grades.
- Have the students provide a small portfolio in which they select two or three entries indicative of their own learning, and attach a short explanation regarding why they chose those particular entries.

Acknowledgments

The authors acknowledge with gratitude the time and effort that the biology faculty contributed to the development and implementation of this new course for prospective elementary and early childhood education teachers. We acknowledge Pam Herndon and Tara Nieves, two students in the experimental biology class, for their instrumental insights and their commitment to improve this course.

This research was supported by a National Science Foundation grant in Teacher Preparation (TIP-915083).

References


