

Beyond Outcomes: Standardizing Performance Assessments In A College Science Course

Laura Henriques <lhenriqu@csulb.edu>
Joanne Olson <jkolson@usc.edu>
Alan Colburn <acolburn@csulb.edu>
Science Education Department, California State University, Long Beach

Paper presented at the annual meeting of
Association for the Education of Teachers in Science, Austin TX, January 1999.

Introduction

California State University, Long Beach, like most other institutions, offers multiple sections of courses for prospective elementary teachers, taught by various faculty members. In order to unify one of these courses while still leaving room for individuality, instructors agreed upon a list of broad required outcomes. Instructors were still free to conduct classes as they deemed appropriate in order to help students achieve the standards. Creating and reaching consensus on the standards was the first step in trying to create equivalence in the multiple sections of a course. The next step was to create appropriate assessments to check students' attainment of the standards. This paper will describe the course in question, the course outcomes and the various assessment strategies used to measure competency. The pilot assessment and the process of refining our procedures to the most recent "final" assessment are described.

Description of Course

A Process Approach to Science, SCED 401, is a class taken by prospective elementary teachers at California State University Long Beach, a highly diverse urban commuter school. All students in the course have taken introductory laboratory classes in biology, geology and physical science. The content of the SCED 401 class is interdisciplinary, pulling from students' previous classroom and lived experiences.

Students earn three semester hours of science credit for this course. The Science Education department offers multiple sections of the course each semester, with 200-250 students enrolling each year. At any time there are at least three instructors teaching 4-5 sections of SCED 401.

We specifically structure this class to help prospective elementary teachers: like science, better understand the nature of science, and develop their abilities at identifying, defining and solving problems the way scientists do.

The course approaches the above goals from multiple directions. The most important thing we do is model exemplary teaching practices for elementary teachers to understand. Students spend the vast majority of class time doing hands-on elementary science activities. Most of the science units students complete require multiple weeks of class time. This provides time for students to explore their thinking about topics, as well as apply new knowledge in different situations.

Activities are open-ended and inquiry based. Curriculum materials are often classic elementary activities, the type of things students are most likely to see and do when they are teachers. During a semester, an observer would see students exploring things like pendulums, mystery powders, batteries & bulbs, and pillbugs. The same observer would also see butterflies hatching, students experimenting with lights and shadows, and cleaning a simulated oil spill. The activities students work on, however, vary somewhat from course section to section—it is the teaching method that is consistent.

Existing Learning Goals

The instructors of the course have developed 12 standards expected of students in all sections of the course. The following are the learning goals as stated in the course syllabus as of fall 1998.

By the end of the semester students will demonstrate proficiency in the following:

1. Generate researchable science questions.
2. Describe or devise procedures to answer scientific questions.
3. Critically examine/critique scientific procedures.
4. Interpret results of scientific investigations.
5. Describe how science differs from other disciplines (nature of science).
6. Demonstrate an understanding of a few basic principles in the life, physical and earth sciences.
7. Describe attributes of elementary teachers who like science.
8. Explain the implications of Piagetian theory to student learning in elementary science.
9. Describe connections between science and everyday life.
10. Explain what "hands-on minds-on" science is.
11. Give an example of how process and content can be taught within hands-on minds-on science activities.
12. Demonstrate e-mail, web browsing & ERIC searching proficiency.

The course's curriculum helps us approach these broad goals. Most of the course's units include companion readings or assignments relating the unit to important science education ideas. Most readings are from NSTA publications and popular Paper presented at the National Conference of the Association for the Education of Teachers of Science, Austin, TX, January 14-17, 1999.

media. Thus, early in the course students read about the advantages of hands-on, minds-on science. Course discussion then relates hands-on science to the activities students just finished. Students see that “hands-on” (a term they are familiar with) does not always translate to “minds-on,” and begin understanding the specific things their instructors do to create hands-on, minds-on science units. (See competency #10)

Similarly, students complete a unit introducing elementary school chemistry, while explicitly learning about the nature of science. Eventually, students must try to determine the contents of an unknown white powder. At the same time, though, they read about the nature of science. Being open-ended and investigation oriented, this unit provides many chances for the instructors to point out parallels between the things students are doing and the nature of science. (Bianchini & Colburn, 1998) (See competency #5)

In the same way, after completing a *Floating & Sinking* unit, we introduce students to Piaget’s work and broad distinctions between the concepts of abstract and concrete ideas. Students have an assignment parallel to this unit when they interview elementary aged children regarding their ability to conserve (one of Piaget’s broad thinking structures). We help our students to see that if *their* students lack the ability to conserve some quantities, the students will be unable to acquire a deep understanding of the concepts of density and buoyancy. At the same time this happens, however, class activities and readings help *our* students better understand the same concepts. (See competency #8)

Finally, technology as an aid to science learning is not slighted in this class. Besides weekly use of e-mail, students also leave the class having “surfed” the web during at least two assignments. They also search for science and educational resources using the Educational Resources Instructional Clearinghouse (ERIC). (See competency #12)

Previous Assessment Practices

Research on current assessment practices guide our thinking about assessment (see for example AAHE Assessment Forum, 1998; Hymes, 1991). Even before beginning this project, we made certain assumptions about how students should be assessed in SCED 401. First, the things that are graded should have some intrinsic value; students should learn something by doing the assignment, irrespective of grades. Second, a wide variety of assignments is a better reflection of students’ abilities and development than a small number of similar assignments. Third, clear writing and proper grammar always “count”. Fourth, students should be able to do certain things when they leave this class—eventually specified with the list of course competencies that is the focus of this paper.

Five assignments are particularly important with respect to the course’s broad goals. Although these assignments continue to evolve, all were present before this project began.

First, hands-on minds-on science units often end with performance tasks and oral reports. In class assessments are consistent with modes of instruction and course goals. Science content and ideas about the nature of science are represented on all quizzes. For example, quiz questions include having students build circuits, determine the contents of an unknown mystery powder, and give possible reasons for discrepant data collected in an experiment. (See competencies #5, 6, 10)

Second, students design and conduct experiments outside class twice during the semester. Early in the course they perform consumer product testing. Through controlled experiments, of their own design, they determine which household products are “best.” Reports are written and shared orally. (See competencies #1-4, 9)

Third, later in the course students must also create a researchable question, devise a repeatable procedure addressing the question, collect and analyze data, and (based on their data analysis) continue their study for another “round” of investigation. The second investigation usually involves testing a prediction or explanation students generated during the first investigation. Students eventually write reports and share results with their classmates. These independent investigations count more than other assignments toward the student’s final grade. The investigations also address more of the course’s broad goals than any other assignment. (See competencies #1-4, 9)

Fourth, students conduct Piagetian interviews with elementary aged children. As they analyze results of the interviews, they realize that some science content they need to know as teachers is not intellectually suitable for young children. (See competency #8)

Finally, to help students appreciate the impact of science around them, they read and review articles, science related television shows, and web sites. These reviews are e-mailed to instructors weekly (See competencies #7, 9, 12)

Complete descriptions of assignments, syllabi, schedules and resource lists are on our websites (www.csulb.edu/~acolburn or www.csulb.edu/~lhenriqu).

Creating Summative Assessment

The rationale for creating a shared list of outcomes was to help create continuity and equivalence between the multiple sections offered. The next logical step was to create an end of course assessment package that was aligned with the course goals and instructional strategies. Since it is difficult to accurately assess the course's goals we collaborated to construct summative assessments for the course. The shared outcomes and end of term assessment provide consistency and ensure rigor across sections.

Before beginning this project, we had been interviewing students individually at the end of the course to assess their learning. These interviews were informal, counting little toward a student's grade, and lacking consistency across instructors. Through this project the instructors worked together to create authentic tasks to be completed by all students at the end of the semester. During the spring of 1998 a variety of performance based assessment tasks were piloted during the exit interviews. Results were analyzed to determine the most effective tasks which were then used in the fall 1998 semester. The assessment will eventually be incorporated into the SCED 401 guidelines. Some of the pilot items tested include:

- data sets from experiments which students interpreted;
- flawed descriptions of experimental designs which students corrected;
- videotapes of poor experimental designs which students critiqued;
- using the internet to find resources;
- instructor-created rubrics for each interview question, complete with exemplars for each;
- writing a letter to the instructor describing how the standards had been met;
- performing science activities and describing processes involved; and
- creating concept maps related to major scientific ideas.

Sample Assessment Items and Findings

For the field testing we developed multiple tasks for a single objective. Different instructors tested the different approaches. For example, competency #3, *critically examine/critique scientific procedures*, was tested in three different ways. Students watched and critiqued a videotape of a bad experimental design; they read a description of poorly designed tests and explained why it was poor; or they gave examples of times they designed procedures and discussed methods of making those procedures better. At the end of the semester the instructors met, discussed the relative merits of each approach for each of the competencies. We then went on to test students with a common version of the exit interview this past fall.

Space prohibits discussing how we assessed every competency. Here are some examples, though, of how we assessed specific course outcomes during the 30 minute exit interview.

Proficiency 3: *Critically examine/critique scientific procedures.*

To assess this proficiency, the instructors of the course developed a brief scenario, and acted it out on videotape. The scenario, lasting about two minutes, shows an experiment being done poorly, and involves the testing of paper airplane designs. Eight different paper airplanes are shown sitting on a picnic table at a park on a windy day. The experimenter announces that he is testing which plane has the best design, and to test this, he will see which plane flies the farthest. Each plane is thrown in a different direction, from different starting points, and by different people. The distance each plane travels is then measured using a wrinkled measuring tape or a seven-inch ruler, again by different people and different starting points. The experimenter then announces that "Plane D went the farthest. It's the best design!"

The scenario has design problems ranging from the research question itself to the control of variables and the need for multiple trials. Students are asked to critique the design after viewing the scenario. One of the instructors developed the following rubric to score students' responses:

- 1 - Student fails to see design flaws. A student at this level does not comment on the importance of controlling variables or using multiple trials.
- 2 - Student focuses on a single aspect of the design, such as the fact that there were multiple people throwing and measuring. A student at this level may overlook other variables that were not controlled and usually does not comment on using multiple trials or a clearer research question.
- 3 - Student is able to identify a majority of the variables that were not controlled, and recognizes the need for multiple trials. A student at this level may recognize the need for a clearer research question.

Many activities done in class focus on the control of variables, so it is not surprising that students focus on this aspect of the scenario more than the issue of multiple trials. This fall two instructors recorded specific responses and the percentage of students (N = 74) who identified each design flaw are listed below:

Measurement location & counting	70%
Multiple directions of throw	65%
Wind speed/direction	41%
Multiple people throwing	35%
Didn't see how planes were made	34%
Needed multiple trials	32%
Measuring devices different	30%
Location of throw varied	27%
Height of throw varied	24%
Research question vague	23%
Need more of each plane/sample size	15%
Needed to see how much farther D went	1%

On the whole, students were quite successful identifying the major flaws with the experiment. Almost all of the students identified the problems with the measuring of the distance that the planes traveled. They correctly identified that the starting points were different and that one experimenter couldn't count very well! A majority of the students also identified that the planes needed to be thrown in the same direction.

Surprisingly, only one third of the students identified the need to conduct multiple trials, and only seventeen students commented on the research question and wondered whether the experiment had actually addressed the question. This aspect of experimental design may need to be made more overt during classroom experiences. This is one example of how assessment results are informing our instruction.

It is important to recognize that the video task has limitations. There are so many problems with the experiment, that it is doubtful that any student will be able to identify them all after a single viewing. Furthermore, as students view the scenario, they may be focusing on only one aspect, and fail to see other design problems. For example, some students were able to correctly identify large numbers of uncontrolled variables, but could not remember the research question. It is important to keep these issues in mind when using such a task.

Proficiency #4: *Interpret results of scientific investigations.*

This proficiency is assessed during the exit interview by having students look at data sets. When we assessed this proficiency during the pilot study, one group of students was asked to describe times that they had to interpret data and discuss what they did. Other groups were given data sets to actually interpret. During the pilot study it became clear to us that some students could talk about data interpretation without providing evidence of doing it. For example, students could easily list times when they interpreted data but this lacked any indication that they were able to do so successfully. As a result, we decided that all students would look at data and make conclusions the next semester. During the pilot study some students were given pendulum data and some students were given plant growth data. The pendulum data included length of pendulum, angle of release, mass of bob and the amount of time for 10 complete swings. The plant data sets included four plants in very different settings. One set of data had plants exposed to different amounts of light and the other data set had plants exposed to different colored light. For both sets the number of leaves and height of the plant were given for a two week period.

During the semester students did experiments with pendula to determine which factors influenced the swing. The pendulum data given during the exit interview closely mirrored the data students collected in class. They also did experiments with plants. The plant experiments vary by section and by semester. Some of the experiments that have been done with plants include the effects of oil on plant growth, effects of fertilizer on plant growth, and effects of light on plant growth. While the plant data presented to students was not identical to the data they collected in class it was similar to the types of data they collected.

In the pilot study it was evident that students receiving the pendulum data had an easier time interpreting data. The data format and procedure for gathering data was familiar to them as they had carried out a similar investigation in class. In most cases they referred back to their lab books. Their comments about the data were often prefaced with statements about what they remembered finding out. Their familiarity with the pendulum experiment coupled with their previous findings guided their interpretation of the data. It was not always clear if students were making conclusions based on the data in front of them or if they were forcing the data to fit their conclusions from class.

The plant data, while similar to class experiments, was novel. The students could not look back to their lab books or journals to find the answer. Students given the plant data sets appeared less able at interpreting results of scientific investigations.

Based on the results of the pilot study we decided to give each student two sets of data to interpret - one familiar and one novel. We found that students who were versatile at interpreting results would be able to make conclusions about both data sets while those students who were more transitional in their abilities would only succeed with one set. In spite of explicit instructions by us to use specific trials for making claims, students still made conclusions based on what they remembered from class. When forced to give examples from the data set to support their claims the transitional students tried to force the data to fit their beliefs (even when their beliefs were incorrect). For example, in some instances a student would say that a 0.1 second difference for 10 swings was significant for one situation yet insignificant for another. Students who had high achievement levels for this proficiency were able to make reasonable conclusions for the plants and pendulum, and the conclusions were supported by data, not hunches or recollections.

Proficiency #10: *Explain what "hands-on minds-on" science is.*

SCED 401 is not a methods course. However, course instructors try to model effective science teaching practices. Class time is occasionally spent helping students understand notable aspects of these practices. In addition, students read two or three small articles about hands-on open-ended science teaching early in the course. The readings help students understand what course activities will be like, and why their instructors are teaching with more open-ended methods.

We assess the students' understandings of the idea via questions at the end of course interview. Typically, the instructor/interviewer asks the student "What does the term hands-on minds-on mean to you?" The interview is a particularly good way to assess student understanding of this proficiency because we can ask follow up questions to have students clarify what they say.

For example, if a student says "Hands-on minds-on science is science teaching in which students are actively engaged in their learning." The instructor can respond "Tell me about what you mean when you say 'actively engaged'." Generally speaking, anything students say with educational jargon, or words similar to those in course readings, is followed by the instructor asking for clarification or further explanation. This follow-up aspect of the interview process is what sets it apart from asking the same question on an essay exam.

We kept careful records about what students said during the pilot test of the interview. This made it easy to compare notes at the end of the course, and discuss the sorts of responses we all believed worthy of lower credit, and those worthy of higher credit.

Example low credit responses were those in which students said hands-on minds-on activities were those in which students would "see it, touch it and also have to think." The response reflects little more than rewording the phrase 'hands-on minds-on.' In addition, merely distinguishing activity from passive lecture or text reading also received lower credit. Students often come to class already believing that hands-on activities are beneficial. The course competency is meant to accent the differences between more open-ended inquiry-based activity and classical guided inquiry where every pupil performs the same activity and is expected to generate the same responses.

Students received higher credit on the question for saying that in hands-on minds-on activities they had to figure out a procedure to answer a question and/or interpret resulting data. Some students also went on to discuss coming up with their own questions. They made references to particular course activities and readings, and said things like "In hands-on minds-on activities, the teacher gives students something to work on without giving them the answers. They go on to learn about the concepts *after* experimenting."

Beside this competency, other course proficiencies also lend themselves well to the interview + follow-up question(s) format. These competencies include describing how science differs from other disciplines (where follow up questions can include things like "So, how would you say science differs from art?"), describing attributes of elementary teachers who like science, discussing implications of Piagetian theory to science teaching, and discussing how science process and content can be taught simultaneously (a reference to the California State Science Framework).

Proficiency #12: *Demonstrate e-mail and web browsing proficiency.*

This proficiency is assessed in two ways. First, students are required to establish an e-mail account at the beginning of the semester and submit ten Article Review assignments to the instructor throughout the semester. An Article Review is a short review and reflection of a current science article, one from each of ten sources, some of which are internet sources (see website for specifics).

Instructors devote class time to help students learn how to use the web and how to search the ERIC database. Additionally, the classroom contains five computers with internet access and students make use of them regularly.

Paper presented at the National Conference of the Association for the Education of Teachers of Science, Austin, TX, January 14-17, 1999.

Each exit interview is conducted in the instructor's office. To assess web browsing proficiency, the students use the instructor's computer, which is already turned on with the university home page displayed. The student is given a topic to find on the internet, such as "Find a lesson plan related to rocks and minerals." The student's search strategy is observed. One instructor used the following rubric to assess students' proficiency:

- 1 - Student is unable to use a search engine, and may try to type the term desired on the URL line on the screen. A student at this level may even roam around the displayed home page, and is unable to conduct a successful search.
- 2 - Student does not select an appropriate search engine or narrow the search using provided categories. A student at this level typically uses the URL line on the screen and types a memorized name of a search engine, such as www.yahoo.com. The student will then type the exact words given by the instructor and hit the Search button. Other students at this level may know of sites that have links to other subject-related sites. They will get to that familiar site, and try to find the requested topic from the list of related sites.
- 3 - Student can use a search engine, typically by hitting the Net Search button and may choose one of the categories listed, such as Education. Once within the category, however, the student will type exactly the words they want to find without further narrowing the search. The student will then scroll through the many sites listed, choosing one that is the closest to the desired topic.
- 4 - Student can use a search engine, and recognizes that some are better suited for certain tasks. The student will choose a search engine that will work well with the task at hand, sometimes by examining the categories listed on each site. The student will choose categories, and will continue to choose subcategories, further limiting the search. When typing, a student at this level may use synonyms or more appropriate terms that will yield a more successful search.
- 5 - Student can use a search engine and may use engines not listed under Net Search in Netscape. The student can use search limiting strategies, such as "&" and "Not" to narrow a search. If too many sites are found, the student will go back and rework the search rather than scrolling through countless site listings.

Students are improving in their ability to browse the internet. The majority of students (over 80%) were at levels three or four. They were able to use search engines and many can limit searches by using categories or more appropriate terms.

Students' previous knowledge and experience have a huge impact on how well they demonstrate this proficiency. In general, the students who do least well with this competency have not yet taken the required computer course. We have decided, however, that we can do more to help students increase their proficiency. During the midpoint of the upcoming semester, we will schedule an additional class session in the computer lab. By this time, students who did not previously have e-mail accounts are beginning to use e-mail regularly and are also more prepared to receive instruction on web browsing. It is hoped that the time will be used to provide specific assistance to students who are still struggling in their use of the technology. This is another example of how assessment results inform instructional strategies.

The next step for the instructors is to come up with a common rubric for this competency. The rubric above is one instructor's view of expert searching. There are other interpretations of expert searching though (Colburn, 1999). During the next semester this rubric, along with others will undergo revision as the instructors struggle to define appropriate performance criteria.

Instructors meet regularly to select (or develop) and refine the assessment instruments and scoring rubrics. As one might expect, different competencies lend themselves to different methods of assessment. For example, it makes the most sense to have students demonstrate internet and ERIC searching skills by having them actually carry out a search. Some of the competencies are measured equally as well with interview as with another approach. All else being equal, the most time efficient and cost efficient approach is employed.

Caveats

There were some difficulties with this approach which should be noted. This process requires a huge time commitment. It takes 30-45 minutes per student to carry out the interview, longer than it would take to grade a final paper or project. This does not factor in the time required to analyze the course, develop commonly agreed upon competencies and create, implement and modify assessments.

Paper presented at the National Conference of the Association for the Education of Teachers of Science, Austin, TX, January 14-17, 1999.

If we did not feel that the interviews offered much richer information than alternate methods we would not employ them. Student feedback indicates they appreciate the one-on-one approach and feel that in most cases they can better represent their true levels of understanding. We like the fact that follow up questions can be asked. This enables us to learn more accurately what our students know and can do. Long Beach has many students for whom English is a second language. The interview format is helpful to this population of students. Perhaps more importantly, it is causing instructors to reflect on the alignment between the course's goals, instructional strategies and assessment approaches.

Getting multiple instructors with different philosophical backgrounds to agree on the competencies and assessment strategies proved more difficult than first anticipated. Three of the four instructors had a similar science education background while one instructor had a strong elementary background with little formal science or science education training. This resulted in different understandings about the nature of science which in turn made consistent interpretations of student responses more difficult. The rubric could alleviate some of the problems but that requires instructor buy-in for the rubric.

The logistics of getting all instructors together to work on the competencies and the assessment strategies, share results, and refine the approach for the next semester is difficult. With part time faculty who are not regularly on campus (and usually not at the same time as the other part timers) the task of finding mutually agreeable time to work is hard. We were lucky to have stipends available to us for creating and testing the new assessment strategies. California State University, Long Beach offered internal grants for faculty working on standards-based learning and assessment. Each faculty member involved in the project received a small stipend for their contributions and we were able to purchase supplies for the assessment. We were ready to move forward with a course-wide final assessment without the grant but it happened much faster with the grant monies available.

Having multiple instructors doing exit interviews which require a VCR and television set at the same time is problematic. Our department has one dedicated TV and VCR and we had three different instructors simultaneously needing it. We were able to get additional monitors and VCR's from the audiovisual department but it was an inconvenience. We are looking into the possibility of streaming the video and having it on the web. That the instructors could individually interview students in their office with their own computers. Computers with web connections are already used during the interview so this would not be an added difficulty.

References

- American Association for Higher Education (1998). *AAHE Assessment Forum: 9 Principles of good practice for assessing student learning*. <http://www.aahe.org/princpl.htm>
- Bianchini, J.A., & Colburn, A. (1998). *Inquiry into inquiry: What do students learn about science's nature through inquiry instruction?* Paper presented at the annual meeting the National Association for Research in Science Teaching, San Diego, CA.
- Colburn, A., (1999). *Finding useful resources on the World Wide Web*. Paper presented at the annual meeting of the Association for the Education of Teachers of Science, Austin, TX, Jan 14-17.
- Hymes, D.L. with Chafin, A.E. & Gonder, P. (1991). *The changing face of testing and assessment: Problems and Solutions*. AASA Critical Issues Report. Eric Document Reproduction Services No. 339 726

NOTE: Since presenting the findings of our assessment grant at Paper presented at the annual meeting of AETS in Austin TX we have continued to work on this project. We have used the assessment strategies developed to unify the class and indoctrinate new part-time faculty. Presentations and papers related to this work are outlined in the table below.

On campus	<ul style="list-style-type: none"> • Henriques, L. & Colburn, A. (1998). Assessing Student Outcomes grant info meeting, Oct.5. • Colburn, A. & Henriques, L. (1999). <i>University Standards in Practice</i> presentation for Standards Based Integrated Teacher Education grant faculty inservice Jan. 20. • INVITED PRESENTATION - Colburn & Henriques were invited to present assessment grant research and findings to ~100 CSULB faculty at The Point, Feb. 26, 1999. (Colburn did presentation)
Regionally	<ul style="list-style-type: none"> • Colburn, A. & Henriques, L., (1998). <i>Helping preservice elementary teachers learn</i> Paper presented at the National Conference of the Association for the Education of Teachers of Science, Austin, TX, January 14-17, 1999.

science content & pedagogy simultaneously. Paper presented at the Symposium on University Teaching, San Bernardino, CA, February 21.

- Colburn, A. & Martin, C. (2001). *Reforming a College Science Course via Assessment and Standards* submitted for presentation at California Science Teachers Association annual meeting, Palm Springs, CA October.

Nationally

- Henriques, L. Olson, J., & Colburn, A. (1999). *Beyond outcomes: Standardizing performance assessments in a college science course*. Paper presented at The Association for the Education of Teachers in Science annual meeting, Austin, TX, January 14-17.
- Henriques, L. & Colburn, A. (2000). *Reforming an innovative college science course via assessment and standards*. School Science and Mathematics Conference. Albuquerque, NM, October 20.

Articles

- Henriques, L., Colburn, A., & Olson, J. (submitted 3/00) Using Standards & Performance Assessment To Unify A Multi-Section Science Class. *Journal of Science Teacher Education*. REVISE & RESUBMIT -- in progress.
-