

Preparing Student Teachers for Alternative Assessment in Science

Norman Herr Margaret Holzer
California State University, Northridge

Megan Martin
California Department of Education

Roxie Esterle
California Academy of Mathematics & Science

Connie Sparks
Kennedy High School

Assessment in American Secondary Schools

Educators have developed a myriad of standardized tests to measure student ability and learning, as well as assess the effectiveness of academic programs. Each year, more than 100 million standardized tests are administered to American public school students. The majority of these tests use a multiple-choice format and report scores using norm-referenced or criterion-referenced scales (Haney & Madaus, 1989; Neill & Medina, 1989). As public perception of the quality of education has waned, there has been an increasing demand for accountability and standardized testing (Cooley, 1991). During the past two decades, many national reports have identified and targeted deficiencies in American education (Education Commission of the States, 1978, 1991; National Commission on Excellence in Education, 1983) and prompted a series of reform initiatives, many of which employ standardized exams (Haney & Madaus, 1989) as mechanisms of accountability and agents of change.

During the 1970s and 1980s, a number of states adopted testing as a major element in their educational reform efforts (Council of Chief State School Officers, 1992; Linn, 1993), and by the beginning of the 1990s, proposals for a national system of testing arose. Under the Bush Administration, the Department of Education issued the America 2000 proposal, suggesting that national tests be used to monitor and stimulate educational progress (US Department of Education, 1991). In response to the

growing enthusiasm for accountability and reform, Congress created the National Council on Education Standards and Testing that concluded:

National standards and a system of assessment are desirable mechanisms for raising expectations, revitalizing instruction, and rejuvenating educational reform efforts for all American schools and students. (National Council on Education Standards and Testing, 1992, p. 8)

Assessment has become a major force for implementing educational policy (Petrie, 1987). Although national examinations have not been adopted, numerous states have implemented their own assessment programs, and in the years between 1984 and 1992, the number of states with science achievement tests increased from 12 to 27 (Council of Chief State School Officers, 1992).

Educators and politicians in some states have suggested that traditional standardized tests, with their focus on factual information rather than problem solving, may actually undermine efforts to raise standards. The California State Department of Education went as far as to say:

The current approach to assessment of student achievement which relies on multiple-choice student response must be abandoned because of its deleterious effect on the education process. An assessment system which measures student

This material is based upon work supported in part by a grant from the National Science Foundation California State University Student Teaching Development Project (DOE-9250027) and the W. M. Keck Foundation. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the funding organizations.

achievement on performance-based measures is essential for driving the needed reform toward a thinking curriculum in which students are actively engaged and successful in achieving goals in and beyond high school. (California Department of Education, 1990, p. 17)

This sentiment has been echoed by other boards and commissions, particularly the National Council on Education Standards and Testing (1992) and the Secretary's Commission on Achieving Necessary Skills (US Department of Labor, 1991) which have recommended that assessment mechanisms be more closely aligned to performance standards. These reformers insist that we need to examine student performance on worthy, real-life intellectual tasks, in place of the simplistic multiple-choice substitutes that are the mainstay of most standardized tests. They argue that authentic assessment tasks should require students to effectively use acquired knowledge rather than merely recognize or recall disarticulated facts and figures. Such tests should present students with an array of tasks that reflect the challenges found in good instructional activities such as designing experiments, developing papers, and dealing with reality-based open-ended situations. Authentic assessments provide challenges that reflect the ambiguities of real-life tasks (Champagne, Newell, & Sigrin, 1992; Newman, 1991; Office of Educational Research and Improvement, 1990; Ravitch, 1993; Schnitzer, 1993; Wiggins, 1989, 1993)

By 1992, 16 states were designing, pilot testing, or implementing alternative assessments that included portfolios, open-ended questions, multiple-choice extensions, and performance assessments (Council of Chief State School Officers, 1992). While state education officials offer new means of authentic assessment in an effort to encourage higher-order reasoning and problem solving, it is doubtful that there will be much success unless teachers are adequately prepared to implement similar forms of assessments in their classrooms.

Purpose

As of 1992, the following states had adopted alternative science assessments that included performance-based tasks, multiple-choice extensions, open-ended questions, and/or portfolios--Arizona, California, Connecticut, District of Columbia, Hawaii, Kentucky, Maine, Maryland, Massachusetts, Minnesota, Missouri, New York, North

Carolina, Texas, and West Virginia (Council of Chief State School Officers, 1992). Although many assessment programs are now well-established, there is little, if any, literature describing how teachers are being trained to prepare students for these exams. Unfortunately, preservice teachers receive little training in either traditional or alternative assessment (Stiggins, Conklin, & Bridgeford, 1986). It is unlikely that new authentic assessments, regardless of their quality, will have a measurable influence on the quality of education in America unless classroom instruction is also affected.

To address these concerns, this project sought to: (a) develop a model program for training science student teachers in alternative assessment, (b) assess the influence of this program on the professional development of participants, and (c) determine ways in which such a program could be institutionalized, if warranted. Arrangements were made with the California Department of Education to train student teachers in alternative science assessment and subsequently employ them in the grading of the statewide Golden State Examinations (GSE).

The Golden State Examinations

The Golden State Examinations were established as a voluntary statewide test to identify and recognize the achievement of individual students throughout the state in specific subject areas. In 1993, Golden State Examinations were offered in seven subjects, including chemistry, biology, and coordinated science. The 1993 science exams consisted of two 45-minute sections that employed alternative assessment items. In addition to traditional multiple-choice questions, the first section included multiple-choice extension and open-ended short answer questions. A multiple-choice extension gives students the opportunity to expand upon concepts that have been developed in a series of related multiple-choice questions. When answering open-ended questions, students may develop a wide variety of acceptable responses depending upon how they define and address the problem. The second section included a laboratory performance task that required students to apply scientific principles in a hands-on laboratory investigation.

The Golden State Examinations were designed to support the 1990 California Science Framework by:

Testing concept-based science content, problem-solving abilities, and science thinking

processes It [the GSE] requires students to synthesize information, analyze and interpret data, organize their thoughts logically and communicate them clearly, apply information to new situations, exhibit creativity, and demonstrate the ability to perform laboratory tasks. (California Department of Education, 1993a, b)

Of the 48,188 high school students who took the 1993 biology exam, 2,007 (4%) earned high honors; 5,216 (11%) earned honors; and 10,560 (22%) earned school recognition. Of the 30,429 students who took the 1993 chemistry exam, 1,167 (4%) earned high honors; 1,950 (6%) earned honors; and 6,922 (23%) earned school recognition. Students received recognition for their performance in the form of certificates, diploma seals, and transcript notations. An additional 1,116 tenth-grade students enrolled in integrated science course work participated in the first field test of the coordinated science examination (California Department of Education, personal communication, 1994).

The Project

In Spring, 1993, seven science student teachers from California State University, Northridge, volunteered to participate in the grading of the Golden State Examinations in biology and chemistry. All received training by recognized experts in the field of alternative assessment. They were introduced to concepts and techniques of alternative assessment and given practice analyzing, answering, and assessing answers to multiple-choice extension, open-ended, and performance-based questions. Student-teachers were introduced to the holistic scoring techniques used to grade the Golden State Examinations, and comparisons were made with traditional forms of assessment. Following the training session, the student teachers traveled to the Educational Testing Service offices in Emeryville, California, where they joined a team of experienced science teachers from throughout the state, many of whom had participated in earlier grading sessions.

At the grading, representatives of the GSE introduced standards for the multiple-choice extension, open-ended, and performance-based questions. After reading a sampling of student answers, participants critiqued the standards and engaged in an energetic discussion of how they might be improved. Once final standards were agreed upon, readers scored a series of sample tests. When a supervisor determined that a reader was accurate

and consistent, he or she allowed the reader to start the official grading process. During the 2-day grading, participants frequently collaborated with other readers regarding the evaluation process and became involved in discussions of pertinent issues of curriculum and instruction

Following the grading session, questionnaires and interviews were conducted with those participating in the project. All student teachers stated that they found the grading to be a valuable professional experience that it should be incorporated into teacher training programs. The ability to interact with established professionals was particularly valued as reflected by the following comment:

It was inspiring to work with experienced professionals, each individually a gold mine of information on teaching techniques, methods, and approaches.

Because the student teachers were so positive about the experience and because the GSE grading team was so pleased with their performance, the training and grading experience was incorporated as a permanent feature of the science teacher preparation program. It was decided that the next cadre of student teachers would participate in the grading of the field test of the state's new Coordinated Science Examination in Winter, 1994. Eleven student teachers representing four campuses of the California State University participated in the grading of the Coordinated Science Examination.

Prior to the initial training session, student teachers were given 15 minutes to review the California State Framework's section on energy for grades 7-9 and 40 minutes to write their best examples of multiple-choice, short answer, essay, and performance-based questions at these grade levels. The student teachers then participated in a training session similar to that provided in the pilot study. Subsequently, they assisted in the grading of the Coordinated Science Examination in a session similar to that described above. Following the grading session, student teachers were asked to provide a written evaluation of their own questions (authored before the training session), complete a brief written survey, and participate in an interview.

Program Evaluation

To ensure quality control in the grading process, all readers were evaluated by their table leaders. An analysis

of the reader evaluation forms shows that the group of student teachers who participated in this project were trained to be as fast, accurate, and consistent as the group of experienced teachers who were hand-picked for the grading. Members of the Golden State Examination Test Development Committee were pleased with the performance of the student teachers as reflected in the following statement:

The preservice teachers . . . were a welcome addition to the GSE scoring. They were enthusiastic, conscientious, and open to learning about holistic scoring and performance-based assessment. Their new assessment skills will benefit students, other preservice teachers, as well as veteran teachers they may work with.

The data and comments of committee members showed that it was possible for student teachers to be employed as readers without compromising the quality of the grading.

In order to evaluate the effectiveness of this project in training science student teachers in alternative modes of assessment, questionnaires, interviews, participants' analyses of test questions, and observations by GSE representatives were used. Unless otherwise stated, the data described below was obtained from 15 student teachers (4 from the 1993 project and 11 from the 1994 project) for whom we had a complete set of data.

The response of science student teachers to the project was extremely positive. All those participating described the project as a "valuable professional experience." Eighty-eight percent said they would definitely recommend the program to other student teachers while the remaining 12% said they probably would. Ninety-four percent recommended that this type of program be incorporated into science teacher training programs at other California State University campuses.

All participants believed the project influenced the manner in which they intend to evaluate student performance. Each of the participants stated that as a result of the GSE grading experience, they had decided to incorporate one or more of the following elements in their curricula and testing--open-ended questions, performance-based tasks, multiple-choice extensions, holistic scoring, and questions that assess ability to problem solve, apply major principles and integrate knowledge. All of these assessment procedures are consistent with the recommendations made by the proponents of authentic assessment (Newman 1991; Schnitzer, 1993; Wiggins,

1989, 1993). All of the student teachers participating in the grading thought that science instruction in California would change if more teachers participated in these types of grading sessions. More than half believed that there would be changes in classroom assessment procedures, and the remainder thought that teachers would place a greater emphasis on major concepts and the integration and application of these concepts.

Participants expressed concern that large numbers of those taking the Golden State Examinations were not adequately prepared for this type of assessment. When asked to identify the area of greatest deficiency, all of them specified skills which are required by authentic and performance-based assessments, but not necessarily by traditional tests. Approximately 60% cited student deficiencies in the higher-order thinking skills needed to apply scientific concepts to novel situations. The remainder cited student inability to interpret and/or formulate written responses to the open-ended and performance-based questions posed on the exam. These capacities can not be evaluated by multiple-choice tests.

After participating in the grading, participants were asked to give general recommendations for science teachers. Approximately half stressed the need to provide students with laboratory science experiences in which students learn to develop hypotheses, design experiments, analyze data, generate and support conclusions, and write reports. The other half of the participants stressed the need to provide students with experiences that build skills required by performance-based examinations. In particular, they recommended giving tests that include open-ended and performance-based tasks, and/or giving students specific instruction on how to read, interpret and respond to such tasks. In addition, approximately 40% of all participants emphasized that teachers should demonstrate how scientific concepts apply to real-life problems and how they apply across traditional disciplinary lines.

All of the student teachers stated that participation in the GSE scoring was a valuable professional experience. Eighty-seven percent described opportunities they had for meaningful dialog with other professional educators about curriculum, assessment, motivational strategies, teaching methodologies, placement opportunities, and professional development. Seventy-three percent described the grading experience as an opportunity to develop professional networks and obtain a broader perspective of science education throughout the state.

After two days of reading responses to the GSE's

open-ended questions and performance-based tasks, student teachers in the 1994 project were asked to identify criteria that would indicate test questions that were poorly constructed. Approximately 90% said that a large percentage of vague or off-task responses might indicate that the questions were confusing or ambiguous. When asked to analyze their own questions written prior to the GSE grading experience, approximately 90% determined that one or more were confusing or ambiguous and offered ways in which they could be rewritten to avoid such problems. In addition, approximately half made specific comments about how their own questions could be rewritten to test higher-order thinking skills and assess student understanding of concepts. Although the lack of a suitable control made it impossible to determine causality, an analysis of the interview transcripts suggests that the GSE grading experience helped student teachers become more aware of potential problems or inadequacies with their own test questions.

Conclusions and Recommendations

The purpose of this project was to: (a) develop a model program for training science student teachers in alternative assessment, (b) assess the influence of this program on the professional development of participants, and (c) determine ways in which such a program could be institutionalized, if warranted.

Analyses of the interview and survey data clearly indicated that those participating believed that this was a positive professional experience that provided them with invaluable experience in alternative assessment, as well meaningful dialog with teacher professionals. It is clear that the experience gave student teachers an excellent opportunity to understand the techniques, challenges, and potential benefits and problems of authentic assessment in science. As a direct result of the GSE grading experience, all the participants stated that they planned to incorporate one or more aspects of authentic assessment in their own teaching. If these individuals follow through with their intentions, then their students will be exposed to teacher constructed tests with open-ended questions, performance-based tasks, extended multiple-choice questions, holistic scoring, and questions that assess ability to problem solve, apply major principles, and integrate concepts.

It was the opinion of the participants in this study that other student teachers could benefit from this type of experience. Although many states have implemented or are developing statewide tests that involve essays, open-

ended questions and performance-based tasks, the grading is generally done only by teachers with many years of experience. Although it may be very important to employ experienced teachers in the grading, this project has shown that, following training, student teachers can perform the grading tasks with equal speed, accuracy, and consistency. It has also shown that student teachers desire to employ many of the new assessment ideas introduced by such examinations as they begin their careers as educators. In addition, student teachers clearly benefit from professional dialog with experienced teachers and have the opportunity to develop networking skills and contacts that may prove invaluable for the continuation of their professional development. Because student teachers are at the beginning of their career, they have the opportunity to influence many generations of students, and it is therefore critical that they receive opportunities for professional development such as the one discussed here.

Although statewide alternative assessments are hailed by many educators as providing a new opportunity to stimulate educational reform, they are likely to produce little improvement if not accompanied by changes in classroom instruction and assessment techniques and are likely to be abandoned in periods of fiscal constraint because the costs required to score them. The findings of this study suggest that the incorporation of student teachers in the grading process can address both of these problems. Student teachers volunteered for this project despite a lack of remuneration or course credit and indicated their intention to use alternative assessment techniques in their classrooms. Since this project has been seen as successful by teacher educators, student teachers, and GSE representatives, this type of opportunity will be offered to science student teachers at all of the 20 campuses of the California State University system, and we encourage other teacher training institutions to investigate similar opportunities within their own states.

References

- Bok, D. (1986). Toward higher learning. *Change*, 18(6), 18-27.
- California Department of Education. (1990). *Education summit*. Sacramento, CA: Author.
- California Department of Education. (1993a). *Golden State Examination in biology: A guide for teachers and students*. Sacramento, CA: Author.
- California Department of Education. (1993b). *Golden*

- State Examination in chemistry: A guide for teachers and students.* Sacramento, CA: Author.
- Champagne, A., Newell, B., & Sigrin, T. (1992). Directions for research and development: Alternative methods of assessing scientific literacy. *Journal of Research in Science Teaching*, 29(8), 841-60.
- Cooley, W. (1991). State-wide student assessment. *Educational Measurement: Issues and Practice*, 10(2), 3-15.
- Council of Chief State School Officers. (1992). *State policies on science and mathematics education 1992.* Washington, DC: Author.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438-481.
- Education Commission of the States. (1978). *National assessment of educational progress: A project of the Education Commission of the States.* Washington DC: Department of Health, Education, and Welfare, Education Division.
- Education Commission of the States. (1991). *National assessment of educational progress: A project of the Education Commission of the States.* Washington DC: United States Department of Education.
- Haney, W., & Madaus, G. (1989). Searching for alternatives to standardized tests: Whys, whats, and whithers. *Phi Delta Kappan*, 70(5), 683-867.
- Linn, R. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, 15(1), 1-16.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, DC: United States Department of Education.
- National Council on Education Standards and Testing. (1992). *Raising standards for American education.* Washington, DC: United States Department of Education.
- Neill, D., & Medina, J. (1989). Standardized testing: Harmful to educational health. *Phi Delta Kappan*, 70(5), 688-696.
- Newman, F. (1991). Linking restructuring to authentic student achievement. *Phi Delta Kappan*, 72(6), 458-63.
- Office of Educational Research and Improvement. (1990). *The case for authentic assessment.* Washington, DC: Author.
- Petrie, H. G. (1987). Introduction to evaluation and testing. *Educational Policy*, 1, 175-180.
- Ravitch, D. (1993). Launching a revolution in standards and assessments. *Phi Delta Kappan*, 4(10), 767-772.
- Schnitzer, S. (1993). Designing an authentic assessment. *Educational Leadership*, 50(7), 32-35.
- Stiggins, R., Conklin N., & Bridgeford, N. (1986). Classroom assessment: A key to effective education. *Educational Measurement: Issues and Practice*, 5(2), 5-17.
- United States Department of Education. (1991). *America 2000: An education strategy.* Washington, DC: Author.
- United States Department of Labor. (1992). *What work requires of schools: A SCANS report for America 2000.* Washington, DC: Author.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703-713.
- Wiggins, G. (1991). Standards, not standardization: Evoking quality student work. *Educational Leadership*, 48(5), 18-25.
- Wiggins, G. (1993). Assessment: Authenticity, context, and validity. *Phi Delta Kappan*, 74(3), 200-214.