Technology-Based Assessments of Student Learning

Edys Quellmalz
Center for Technology in Learning
SRI International

The Center for Technology in Learning at SRI International has conducted a number of projects creating assessment designs and examplars relevant to the geosciences. In general our approach has been to focus on assessment designs that can probe more deeply into conceptual understanding and extended inquiry than traditional assessments typically do. Our methods draw from theories of learning to develop cognitively principled assessment designs. This paper summarizes the design principles we have forged and some of the assessments we have developed that are particularly relevant to geoscience.

Assessment Design Principles. Research in cognitive science on the development of expertise in many domains indicates that individuals proficient in a domain have large, organized, interconnected knowledge structures and well-honed domain-specific problem-solving strategies (Bransford et al., 2000). Balanced assessment systems, therefore, should aim to measure both the extent and connectivity of students' growing knowledge structures and their problem-solving strategies (Pellegrino et al., 2001; Glaser, 1991). Many assessments in science tend to focus on conceptual understanding, often using multiple choice formats. Although inquiry skills are important standards in all sciences, they are not easily measured by traditional formats, and they are less frequently addressed in formative or summative achievement tests.

Much of our work at SRI's Center for Technology in Learning has employed performance assessments to elicit evidence of students' ability to go beyond selections of correct answers in highly-structured problems. We have designed performance assessments in which students solve authentic, complex problems requiring use of key concepts and use of inquiry strategies. Performance assessments are particularly well suited to measuring students' conceptual understandings and abilities to conduct and communicate investigations of significant, recurring problems (Baxter & Glaser, 1998; Bransford et al., 2000; Pellegrino et al., 2001). Quellmalz and Haydel (2003) also found in cognitive analyses of think-alouds that students were more likely to use schematic and strategic knowledge on performance assessments than on multiple-choice items. Assessment approaches that require students to construct and explain thinking as they solve problems can measure distinct components of inquiry and problem solving, including stating research questions, posing hypotheses, planning and conducting investigations, gathering evidence, analyzing data, considering disconfirming evidence, and communicating explanations.

Our assessment designs incorporate advances in measurement science that integrate cognitive research findings into systematic test design frameworks. Evidence-centered assessment design is a method for structuring an assessment argument by relating the learning to be assessed, as specified in a student model, to a task model that specifies features of tasks and questions that would elicit the evidence of the learning targets, then to an evidence model that specifies the quality of student responses that would indicate levels of proficiency (Messick, 1994; Mislevy et al., 2003; Pellegrino et al., 2001). Simply put, principled assessment design involves specifying the knowledge and skills to be tested, the types of tasks and items that would elicit evidence of the learning, and the scoring that would report levels of progress.

We have also created a modular approach to the design of extended performance assessments. In these designs, a unifying problem or driving question is addressed in a series of related modules that focus on skill sets related to planning, designing, and conducting an investigation, observing and displaying data, analyzing and interpreting data, drawing conclusions, employing alternative representational formats, and communicating results in the form of a scientific argument. The modules can vary according to the complexity of the content, inquiry, and technology required. Thus a template for an assessment can shape numerous variations of component modules to fit the assessment purpose and population.

Assessment Resources. Types assessment resources we have developed include:

- Digital collections of performance assessments for science, mathematics, and technology;
- Alignment protocols and tools for linking assessments to standards and curricula;
- Online professional development tools for creating and adapting performance assessments;
- Online tools for professional development on scoring student work;
- Prototype assessment exemplars for assessing science inquiry, mathematical problem solving, data literacy, and use of a range of technology tools including visualizations and modeling tools.

Digital Collections. We have developed online collections for science, mathematics, and technology.

PALS. Performance Assessment Links in Science (PALS) is a well-established digital library of resources and technical assistance that supports science and assessment reform that can "break the mold" of on-demand, traditional assessment (see: http://pals.sri.com/). PALS, an online, standards-based, interactive resource bank of more than 300 K-12 science performance assessments with documented technical quality, pioneered digital library collections and assessment resources (Quellmalz, Schank, Hinojosa, & Padilla, 1999). Assessments are indexed to science and mathematics standards and to curriculum programs.

PALS differs from other assessment collections in that it draws assessment that have been developed by a wide range of established assessment development programs. The system can be searched for assessments by science and mathematics standards, curriculum unit, grade range, and content area. The system generates search results as assessment planning charts (Stiggins, Rubel, & Quellmalz, 1986), showing which tasks in the collection are designed to test selected standards. A PALS Guide offers professional development guidelines for adapting and developing science assessments (Stiggins, 2002). Communities of practice are supported by task rating and comment tools in a threaded discussion board and through informal and formal sessions in the CTL virtual professional development center, TAPPED IN™ (Schlager, Fusco, & Schank, 1998; http://tappedin.sri.com/).

Visits to PALS have exceeded150,000 in total, averaging 20,000 visits per month. PALS users include teachers, teacher professional development programs, assessment programs, curriculum evaluators, and researchers. The PALS Web site is linked to the National Science Teacher Association's SciLinks, the Eisenhower National Clearinghouse, ERIC, and the U.S. Department of Education's Gateway to Educational Materials Web site, among others.

Data evaluating PALS have been collected from a wide variety of users. Teachers and professional developers show great enthusiasm for PALS resources. Respondents to evaluation questionnaires indicate that PALS resources were highly useful and easy to navigate. A study of two district models for using PALS indicated that district personnel and teachers regarded the PALS resources as the only source of high-quality and useful assessments of standards. Moreover, many agencies and organizations have written letters of support for the PALS project and its further development. (Quellmalz, 2003).

PALM. A sister site under development, Performance Assessment Links in Mathematics (PALM), can be viewed at http://palm.sri.com . The site presents both performance assessments for mathematics and ones developed to assess science but that require use of mathematics. The tasks have been aligned with NCTM and with NSES.

IPAT. The Integrated Performance Assessments of Technology (IPAT) Web site presents our innovative assessments of students' abilities to use technology to solve complex problems. http://ipat.sri.com These assessments illustrate the modular design approach. For example, in one assessment students are use the ArcView visualization tool to gather data to determine which states have appropriate climate conditions that would allow them to apply for solar energy grants.

Innovative Assessment Designs. We have developed a number of assessment tools and prototypes that test students' abilities to use science knowledge and technologies to solve significant, recurring, authentic science problems.

GLOBE. In the Global Learning Observations to Benefit the Environment (GLOBE) project, we developed classroom assessment tools that test students' deep understanding of GLOBE concepts and their ability to conduct and interpret GLOBE environmental investigations. For example, in GLOBE learning activities, students learned about visualization of climate data, phenology, and the reasons for seasonal change. We indexed the GLOBE assessment framework to the National Science Education Standards (NSES), the Third International Mathematics and Science Study (TIMSS) science framework, and the National Assessment of Educational Progress (NAEP) science framework. Assessment tools included tests of GLOBE students' appropriate use of measurement protocols and solutions to integrated investigation problems. These assessments take data from the GLOBE data archives and have students solve authentic problems, analyze and interpret the GLOBE data, and communicate their findings and recommendations. Templates for the assessments serve as models for teachers to develop new assessments. In addition, the project has developed approaches for aligning GLOBE with state standards (see http://globeassessment.sri.com). SRI and GLOBE systems staff designed an alignment database linking state science standards to GLOBE materials.

Calipers. One of our most recent NSF projects is Calipers: Using Simulations to Assess Complex Science. We are partnering with Concord Consortium in a project to demonstrate the value of simulations for measuring deep conceptual understanding and extended inquiry. The project will document the prototype assessments technical qualities, particularly their validity. We will also examine the logistical and economic advantages on technology-based performance assessments.