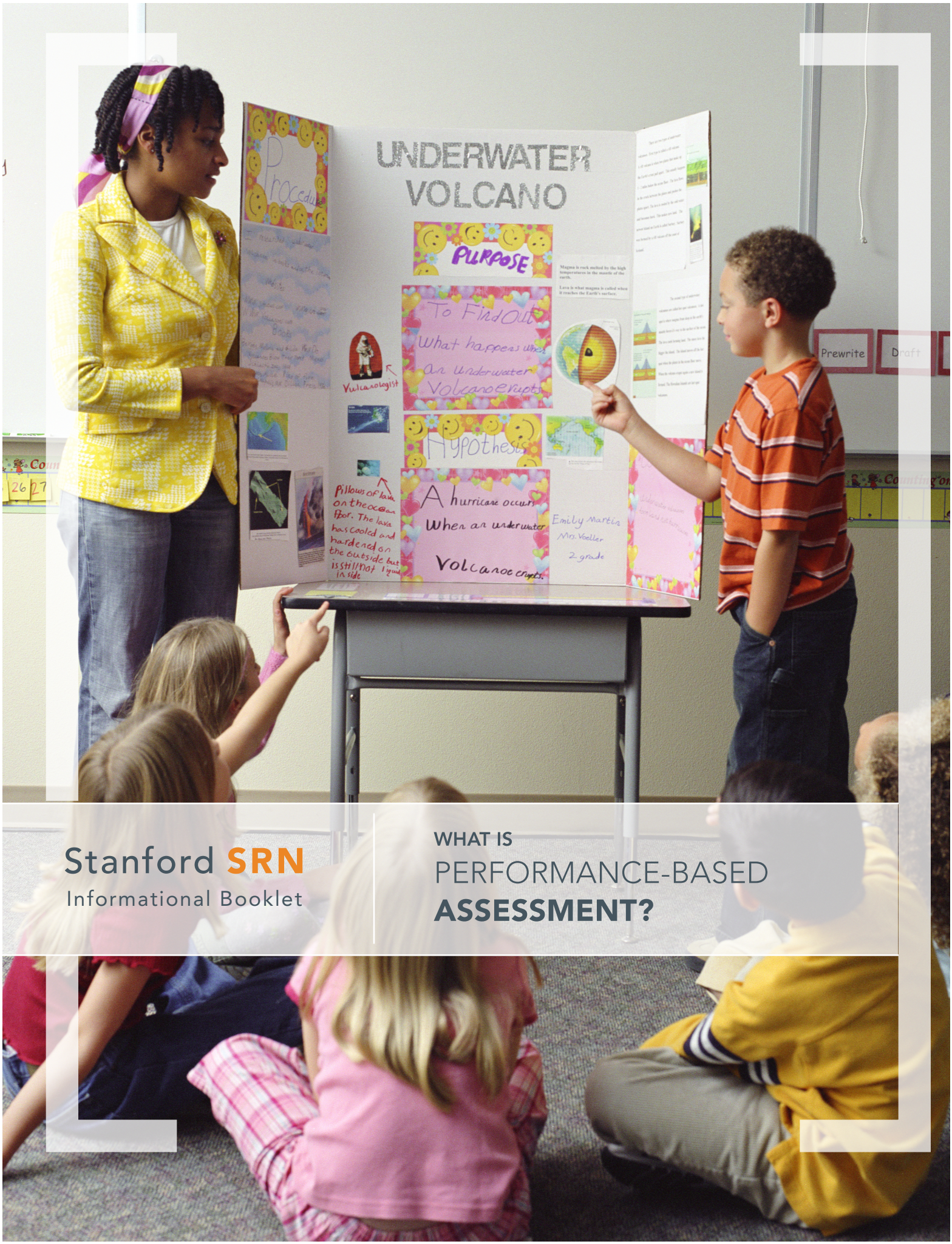


Stanford **SRN**

Informational Booklet

WHAT IS PERFORMANCE-BASED **ASSESSMENT?**

WHY PERFORMANCE-BASED ASSESSMENT?

As the vast body of knowledge continues to expand, it is becoming impossible for individuals to keep up with the amount of information available even in a single field. This trend, along with technological progress, has transformed the labor demands of the world economy.

According to the New Commission on the Skills of the American Workforce (2007):

> The core problem is that our education and training systems were built for another era, an era in which most workers needed only rudimentary education…. [The world of the future is] a world in which routine work is largely done by machines…in which line workers who cannot contribute to the design of the products they are fabricating may be as obsolete as the last model of that product. (p. 7)

These economic trends and the training needed for the new workforce require that school systems shift from a fact-oriented curriculum to one that emphasizes problem solving and innovation (Herman, 1992). This approach is at odds with the current state of the U.S. education system.

A growing number of business and education leaders also have begun to recognize the importance of the kinds of assessments that are used to evaluate student learning. The Partnership for 21st Century Skills (Fadel, Honey, & Pasnik, 2007, p. 34), for example, has suggested that in an age of innovation—in which the workplace will require "new ways to get work done, solve problems, or create new knowledge"—the assessment of students will need to be largely performance based so that students can show how well they are able to apply content knowledge to critical thinking, problem solving, and analytical tasks throughout their education. Likewise, in *College Knowledge*, author David Conley (2005) reports that higher education faculty value "habits of mind"—including the ability to think critically and analytically, to independently draw inferences and reach conclusions, and to solve problems—even more than they value content knowledge.

More than standardized tests of content knowledge, performance-based tasks are able to measure students' habits of mind. Performance-based assessment requires students to use high-level thinking to perform, create, or produce something with transferable real-world application. Research has shown that such assessment provides useful information about student performance to students, parents, teachers, principals, and policymakers. Research on thinking and learning processes also shows that performance-based assessment propels the education system in a direction that corresponds with how individuals actually learn.

PRACTITIONER SUPPORT FOR PERFORMANCE-BASED ASSESSMENT

Educators who have worked in systems that use performance-based assessment report that such assessment has a positive impact on instructional practice and provides valuable information. In a study of the Kentucky Instructional Results System (KIRIS), which assessed student progress through a combination of open-ended response items, multiple-choice items, portfolios, and performance events, almost 90 percent of principals and 77 percent of teachers reported that the performance assessment was useful for judging the effectiveness of schools. Even more important, performance assessment contributed to improved instructional practices: "40 percent of teachers reported that the open-response items and portfolios have a great deal of positive effect on instruction, and virtually none reported that about multiple-choice items" (Matthews, 1995, p. 11). A report on the Maryland School Performance Assessment Program (MSPAP), another performance-based assessment program, similarly found that "98 percent of school principals felt MSPAP has a positive effect on instruction" (Koretz, Mitchell, Barron, & Keith, 1996, p. 29). Recent experience makes it clear that performance-based assessment provides a means to assess higher-order thinking skills and helps teachers and principals support students in developing a deeper understanding of content (Vogler, 2002).

RESEARCH SUPPORT FOR PERFORMANCE-BASED ASSESSMENT

Research over the years has shown that how student learning is assessed can play an important role in a student's overall learning. As Resnick and Klopfer (1989) point out, content and process are inextricably linked. This connection makes it extremely important to assess students in meaningful ways to determine if they are mastering the content (Herman, 1992). Cognitive psychologists studying how individuals learn have come to the following understanding:

> Mere acquisition of knowledge and skills does not make people into competent thinkers or problem solvers. To know something is not just to passively receive information, but to interpret it and incorporate it; meaningful learning is reflective, constructive and self-regulated (Wittrock, 1991, Bransford and Vye, 1989, Marzano et al., 1988, Davis et al., 1990). (Herman, 1992, p. 15)

An exclusive reliance on multiple-choice tests that primarily measure basic skills and discrete knowledge—but neglect complex thinking and problem solving—is not consistent with what practitioners in the field know about the kinds of assessments that promote student learning.

Performance-based assessment is consistent with modern learning theories and also helps teachers employ what the profession considers to be best practices. The MSPAP report (Koretz et al., 1996) found that in implementing performance-based assessment, teachers changed their instructional practice to do the following: emphasize cooperative work; focus more on writing, problem solving, and real-world, hands-on activities; and deemphasize rote learning and teaching. A study of the Massachusetts Comprehensive Assessment System found that after moving to a performance assessment, instructional practices began to correspond more with those deemed as best practices (Vogler, 2002). Teachers also reported that the open discussion of performance standards and the professional development received regarding scoring performance tasks were powerful professional development experiences.

PERFORMANCE ASSESSMENT: THE NEXT GENERATION OF ASSESSMENT FOR MEETING THE DEMANDS OF THE 21ST CENTURY

Performance assessment uses tasks that require students to demonstrate their knowledge, skills, and strategies by creating a response or a product (Rudner & Boston, 1994; Wiggins, 1989). Unlike a traditional standardized test in which students select one of the responses provided, a performance assessment requires students to perform a task or generate their own responses. For example, a performance task in writing would require students to actually produce a piece of writing rather than answering a series of multiple-choice questions about grammar or the structure of a paragraph. Performance assessment is authentic when it mimics the kind of work that is done in real-world contexts. For example, an authentic performance task in environmental science might require a student to conduct research on the impacts of fertilizer on local groundwater and then report the results to the public through a public service announcement or informational brochure.

Performance assessment taps into students' higher-order thinking skills, such as evaluating the reliability of sources of information, synthesizing information to draw conclusions, or using deductive/inductive reasoning to solve a problem. Performance tasks may require students to make an argument with supporting evidence in English or history or social science, conduct a controlled experiment in science, or solve a complex problem or build a model in mathematics. Performance tasks often have more than one acceptable solution or answer and also require students to explain their reasoning. The format of performance assessment may range from "on-demand" kinds of tasks that can be completed in a given amount of time (a timed writing exercise, for example) to long-term projects that involve independent work or research outside of class.

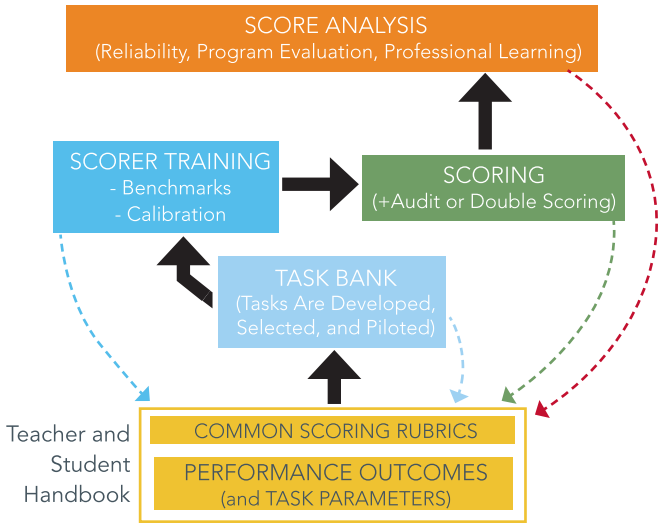
Performance assessment typically is evaluated using rubrics. Rubrics display indicators of performance on the selected evaluative criteria across a range of levels. These levels are written to represent the appropriate range of student performance (such as lower elementary, middle school, or high school). Some rubrics are designed to score a performance task holistically, while analytic rubrics are designed to be scored across multiple dimensions to represent the work in a fine-grained way that allows for more specific feedback to students and instructors.

Performance assessment is used for both formative and summative purposes. When students are provided with multiple opportunities to learn and apply the skills being measured and opportunities to revise their work, performance assessment can be used to build students' skills and also to inform teachers' instructional decisions. When combined with other kinds of assessments of student learning as part of a multiple-measures assessment system, performance assessment can be used for summative judgments about students' understandings and skills in particular domains.

BUILDING A CREDIBLE AND DEFENSIBLE **PERFORMANCE ASSESSMENT SYSTEM**

The following steps can be used to build a performance assessment system, as shown in Figure 1.

Figure 1. Components of a Performance Assessment System



1. For each content area/discipline, the first step is to define the **performance outcomes**, or learning targets, that the performance tasks will assess. The performance outcomes serve as the foundation for the development of the scoring rubrics and performance tasks. To ensure content validity, performance outcomes are aligned with state or national standards, college readiness standards, and the core skills of the discipline. Teachers and other stakeholders—in addition to content-area specialists, assessment specialists, and higher education faculty—are included in the development process to ensure that the performance outcomes reflect the values and priorities of the users of the assessment system.

2. Based on the performance outcomes, **task parameters** (or "task shell") are defined to ensure that the designed performance tasks will measure the desired outcomes. Key decisions are made about task design by answering the following questions:
   - What is the genre of performance that we want to measure (e.g., a scientific inquiry, a literary analysis, or a mathematics application)?
   - How will students communicate their learning (e.g., through a research paper, a lab report, or a multimedia product)?
   - Will the task require independent research, or can it be completed with resources provided in class?
   - How much choice will teachers and students have in determining the content of the performance task?

3. Next is the development of the **common scoring rubrics** that will be used to assess the student work. These rubrics are aligned with the performance outcomes and are organized to represent key dimensions of performance. Written to reflect students' developmental trajectories, rubric levels make clear distinctions between levels to facilitate reliable scoring. The common scoring rubrics are not task-specific and can be applied to evaluate any tasks that are designed to meet the performance outcomes within disciplines.

4. Content-specific **performance tasks** are designed using a backward-planning tool to ensure alignment with the performance outcomes, specific content standards, or other learning targets. The designed tasks are vetted by content-area experts, assessment specialists, and other stakeholders (e.g., teachers or higher education faculty), and approved tasks are entered into the **task bank**. The performance tasks are piloted, and student work samples are produced.

5. After collecting student work samples from across school sites, **benchmark** work samples (those representing different levels of performance on the rubric) are selected for training purposes. Teachers are trained to score student work with the common scoring rubrics. A common training module and scoring procedures maximize score reliability and comparability across schools. Trained scorers then independently score several prescored tasks to check their ability to score reliably. Those who pass the standard for scoring accurately are considered reliable scorers (**calibrated**).

6. Teachers and other participants who have been trained to score and are calibrated in a particular content area participate in scoring the student performance tasks in that content area. These scores are collected and analyzed. The results inform program review and instructional practice as well as provide the basis for further revisions of the performance outcomes, rubrics, and tasks.

7. To check on score reliability and the comparability of scores across teachers and schools, two strategies may be followed: An independent external audit of local school scores may be conducted, or some percentage of student work may be double-scored at the school site. A combination of these two methods may be used to check score reliability within and across schools.

8. Additional research is conducted on students' performance task scores and work samples to evaluate the following:
   - **Content validity** (whether these work samples truly measure state content standards or represent college readiness skills)
   - **Concurrent validity** (how consistent students' performance tasks scores are with students' grades, SAT scores, or other standardized test scores)
   - **Predictive validity** (how well students' performance task scores predict performance in college)
   - **Consequential validity** (what students learn from completing a performance task, or what teachers learn from implementing these tasks)

EXAMPLES OF PERFORMANCE ASSESSMENT SYSTEMS

The following examples represent several assessment systems in local, national, and international contexts that currently implement performance assessment as one measure of student competencies and/or school performance.

INTERNATIONAL BACCALAUREATE DIPLOMA PROGRAM

The International Baccalaureate (IB) Diploma Program enrolls more than 650,000 students worldwide who work toward an IB diploma. The program, designed for students in Grades 11 and 12, assesses students using school-based assessments throughout the two-year program and external exams at the end of the program. Both types of assessments measure students' individual performance on the objectives outlined by syllabi, or "subject outlines," written by the International Baccalaureate Organization.

School-based assessments contribute between 20 percent and 30 percent of the student's total grade—with the exception of arts courses such as music, theater arts, and visual arts, which have assessment components that account for as much as 50 percent of the total grade. Examples of coursework that teachers might grade include oral exercises in language subjects, projects, student portfolios, class presentations, practical laboratory work, mathematical investigations, and artistic performances.

The external exams usually consist of essays, structured problems, short-response questions, data-response questions, text-response questions, case-study questions, and a limited use of multiple-choice questions. External assessment tasks are designed, administered, and graded by IB examiners. The International Baccalaureate Diploma Program exists in public schools throughout the United States and in schools around the world. States such as Colorado and Texas enacted laws that give students college credit for the successful completion of their IB program.

ENVISION SCHOOLS GRADUATION PORTFOLIO

Envision Schools, a public charter management organization that currently runs four small high schools in the San Francisco Bay Area, partnered with the Stanford School Redesign Network to develop a collection of performance tasks across six content areas: English language arts, history/social science, mathematics, science, world language, and creative expression. Teachers collaborate within and across disciplines to design and implement performance tasks within the context of their course curriculum.

Students' performance tasks are scored using subject-specific rubrics that are designed to capture the core skills and understandings of the particular task genre (e.g., scientific inquiry or literary analysis) and to assess students' readiness for college-level work and the workplace. If a student's work is scored at the *proficient* level, it is considered "certified" and can be selected by students as an entry in their Graduation Portfolio. To pass the Graduation Portfolio, students must include certified work in each content area, plus a number of other required artifacts. In an oral defense at the end of the senior year, students present their portfolios to a panel of teachers, advisors, and community members and justify their readiness to graduate from high school. The Graduation Portfolio is an addition to the California graduation requirements. (Students also must earn a minimum number of credits and pass the California High School Exit Exam.) To ensure credibility of the scoring and "certification" process, Envision regularly brings teachers together within and across the four schools to participate in scorer training, calibration, and professional development activities that involve teachers in evaluation of student work. In the future, Web-based scoring tools will make it possible to check how well teachers are calibrated to the scoring standard across content areas and schools.

VERMONT MATHEMATICS AND WRITING PORTFOLIOS

Vermont's Department of Education supports schools and districts that use portfolios for local assessment purposes in all grades. Portfolios typically contain examples of students' best work, with scores reflecting optimum performance. Teachers maintain autonomy in assigning problems for the portfolios and determine the rules for producing, reviewing, and revising student work.

A sample of mathematics and writing portfolios is submitted by each school to a regional meeting for scoring by other teachers. Each portfolio task is scored across seven dimensions on a four-point scale. Composite scores are calculated, but total scores are not reported. In the writing portfolios, each student identifies his or her "best piece," which is scored on five dimensions, with the remaining pieces scored as a set. Each writing and mathematics task is scored twice to assess score reliability. The scores are used primarily to inform instruction and not for accountability purposes.

QUEENSLAND, AUSTRALIA, SCHOOL-BASED PERFORMANCE ASSESSMENT

As one of the two highest achieving states in Australia, Queensland has the most highly developed system of local performance assessment. In Queensland, school-based assessments are developed, administered, and scored by teachers in relation to the national curriculum guidelines and state syllabi (also developed by teachers) and are moderated by panels that include teachers from other schools as well as at least one professor from the tertiary education system.

At the end of the year, teachers collect a portfolio of each student's work, which includes subject-specific assessment tasks, and grade it on a five-point grading scale. To calibrate these grades, teachers put together a selection of portfolios from each grade level—one from each of the five score levels plus borderline cases—and send these to a regional panel for moderation. The panel of five teachers rescores the portfolios and confers about whether the grade is warranted, making a judgment on the range of scores. A state panel also looks at work samples across schools. Based on these moderation processes, the school is given instructions to move grades up or down so that they are comparable to others.

REFERENCES

Conley, D. T. (2005). *College knowledge: What it really takes for students to succeed and what we can do to get them ready.* San Francisco: Jossey-Bass.

Fadel, C., Honey, M., & Pasnik, S. (2007, May 18). Assessment in the age of innovation. *Education Week*, 26(38), 34, 40. Retrieved August 26, 2008, from http://www.edweek.org/ew/articles/2007/05/23/38fadel.h26.html

Herman, J. L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association for Supervision and Curriculum Development.

Koretz, D. M., Mitchell, K., Barron, S. I., & Keith, S. (1996). *The perceived effects of the Maryland School Performance Assessment Program: Final report* (CSE Technical Report No. 409). Los Angeles: University of California, Center for the Study of Evaluation.

Matthews, B. (1995). *The implementation of performance assessment in Kentucky classrooms.* Louisville, KY: University of Louisville.

New Commission on the Skills of the American Workforce. (2007). *Tough choices or tough times.* Rochester, NY: National Center on Education and the Economy.

Resnick, L., & Klopfer, L. (Eds.). (1989). *Toward the thinking curriculum: Current cognitive research.* Alexandria, VA: Association for Supervision and Curriculum Development.

Rudner, L. M., & Boston, C. (1994). Performance assessment. *ERIC Review, 3*(1), 2–12.

Vogler, K. E. (2002). The impact of high-stakes, state-mandated student performance assessment on teachers' instructional practices. *Education, 123*(1), 39.

Wiggins, G. (1989, May). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, 70*, 9.

ADDITIONAL SOURCES

Darling-Hammond, L., & McCloskey, L. (2008). *Benchmarking learning systems: Student performance assessment in international context.* Stanford, CA: Stanford University.

Koretz, D., Stecher, B., Klein, S., McCaffrey, D., & Deibert, E. (1993). *Can portfolios assess student performance and influence instruction? The 1991–92 Vermont experience.* Washington, DC: RAND Institute on Education and Training; and Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Stecher, B. (1998). The local benefits and burdens of large-scale portfolio assessment. *Assessment in Education, 5*(3), 335.

SRN LEADS
STANFORD UNIVERSITY

THE SCHOOL REDESIGN NETWORK AT STANFORD UNIVERSITY

The School Redesign Network (SRN) was established in 2000 at Stanford University to build, capture, and share research-based knowledge to transform secondary schools and school systems. Our mission is to help support and sustain equitable schools and districts that are intellectually rigorous, high performing, and designed to help all students master the knowledge and skills needed for success in college, career, and citizenship.

SRN provides technical consulting and support to schools and districts that have committed to adopting performance-based assessment as part of a multiple-measures system for evaluating student learning and measuring school performance. SRN contracts with schools and districts to develop assessment materials, establish and oversee scoring procedures, provide training and professional development to support teachers engaged in the work, and conduct research to support the validity and reliability of the assessment system. One of the core principles that guide our work is that a performance assessment system should be designed to be educative for students, teachers, and schools.

SRN's Co-Executive Director, Dr. Raymond Pecheone, has more than 30 years of expertise and experience with performance assessment. Among his many assessment-related projects, Dr. Pecheone was involved in the New Performance Assessment Consortium in the San Francisco Bay Area, the redesign of the New York State Regents examinations, and the design of the National Board for Professional Teaching Standards (NBPTS) Portfolio. He led the development and implementation of the Connecticut Beginning Educator Support and Training (BEST) Portfolio Assessment as well as the Performance Assessment for California Teachers (PACT).

Additional information about SRN can be found at **www.srnleads.org**. For more information related to performance assessment initiatives, contact Ruth Chung Wei, Associate Director for Research, Development, and Policy (rchung@stanford.edu; 650-723-8399) or Linda Carstens, Director of Professional Learning (carstens@stanford.edu; 650-736-1529).

School Redesign Network • Stanford University School of Education • 505 Lasuen Mall, Stanford, CA 94305

ACKNOWLEDGMENTS

The School Redesign Network at Stanford University acknowledges the support of the following organizations in the production of this brochure:

Ohio Department of Education

REL MIDWEST
Regional Educational Laboratory

3090_09/08