An Unfinished Canvas




A Review of Large-Scale Assessment in K–12 Arts Education

Research conducted by SRI International

This study was commissioned by The William and Flora Hewlett Foundation.

Copies of this and other *An Unfinished Canvas* reports can be downloaded from
http://policyweb.sri.com/cep/publications

Suggested citation:
Stites, R. & Malin, H. (2008). *An unfinished canvas. A review of large-scale assessment in K–12 arts education.* Menlo Park, CA: SRI International.

Silicon Valley-based SRI International is one of the world's leading independent research and technology development organizations. Founded as Stanford Research Institute in 1946, SRI has been meeting the strategic needs of clients for more than 60 years. SRI's Center for Education Policy studies reforms that hold promise for improving the K-16 system of schooling and lifelong learning. The Center conducts research and evaluations on the design, implementation, and impact of educational programs, especially improvement efforts. For more information about SRI, please visit our Web site: www.sri.com.

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
Phone: 650.859.2000

An Unfinished Canvas

A Review of Large-Scale Assessment in K–12 Arts Education

Regie Stites

Heather Malin

Center for Education Policy

SRI International

2008

CONTENTS

EXHIBITS

ACKNOWLEDGEMENTS

We thank Kris Alexander of The California Arts Project, Nancy Pistone, arts consultant to the Ohio Department of Education, and Lois Hetland of Project Zero at Harvard University for their careful reading and helpful suggestions on a draft of this review. We also thank Tara Pearsall of the South Carolina Arts Assessment Program, Pamela Paulson of the Perpich Center for Arts Education in Minnesota, Dennis Horn of the Collaborative for Teaching and Learning, Phil Shepherd of the Kentucky Department of Education, and David Thurmond of the Oldham County Arts Center for assistance in answering questions about particular arts assessments. We express appreciation to SRI International colleagues who helped in the production of this review. Katrina Woodworth provided leadership and sound advice at many critical stages in the work on this review. Marilyn Gillespie made very useful suggestions for improving the organization of content in an early draft of the review. For editorial assistance, we thank Mimi Campbell; for graphics and production assistance, we thank Kathy Wright and Cher Cai.

Finally, we are grateful to The William and Flora Hewlett Foundation for sponsoring this research. Any opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the reviewers or The William and Flora Hewlett Foundation.

PROLOGUE: KEY FINDINGS AND RECOMMENDATIONS FROM *AN UNFINISHED CANVAS*

In early 2007, SRI International published *An Unfinished Canvas. Arts Education in California: Taking Stock of Policy and Practice*, a statewide study on the status of arts education in California. That study's findings served as the impetus for a series of follow-up studies, including this review of current models for large-scale arts assessment and state arts assessment systems. A summary of key findings from *An Unfinished Canvas* follows.

KEY FINDINGS

Overview of Arts Education in California

- 89% of California K-12 schools fail to offer a standards-based course of study in all four disciplines—music, visual arts, theatre, and dance—and thus fall short of state goals for arts education.
- Methods of delivering arts instruction vary by school level, often resulting in a limited experience at the elementary level and limited participation at the secondary level.
- 61% of schools do not have even one full-time-equivalent arts specialist, although secondary schools are much more likely than elementary schools to employ specialists.
- At the elementary level, arts instruction is often left to regular classroom teachers, who rarely have adequate training.
- Arts facilities and materials are lacking in most schools.
- Standards alignment, assessment, and accountability practices are uneven in arts education and often not present at all.

Arts Education in Elementary Schools

- 90% of elementary schools fail to provide a standards-aligned course of study across all four arts disciplines.
- Elementary students who receive arts education in California typically have a limited, less substantial experience than their peers across the country.
- Inadequate elementary arts education provides a weak foundation for more advanced arts courses in the upper grades.

Arts Education in Middle and High Schools

- 96% of California middle schools and 72% of high schools fail to offer standards-aligned courses of study in all four arts disciplines.
- Secondary arts education is more intense and substantial than elementary arts education, but participation is limited.

Change over Time in Arts Enrollment

- Enrollment in arts courses has remained stable over the last 5 years, with the exception of music, which has seen a dramatic decline.

Unequal Access to Arts Education

- Students attending high-poverty schools have less access to arts instruction than their peers in more affluent communities.

Barriers to Meeting the State's Arts Education Goals

- Inadequate state funding for education is a top barrier to the provision of arts education, and reliance on outside funding sources, such as parent groups, creates inequities.
- Pressure to improve test scores in other content areas is another top barrier to arts education.
- At the elementary level, lack of instructional time, arts expertise, and materials are also significant barriers to arts education.

Sources of Support for Arts Education

- Districts and counties can play a strong role in arts education, but few do.
- Schools are increasingly partnering with external organizations, but few partnerships result in increased school capacity to provide sequential, standards-based arts instruction.

RECOMMENDATIONS

State Policymakers

- Increase and stabilize education funding so that districts can develop and support a standards-based course of study in each of the four arts disciplines.
- Strengthen accountability in arts education by requiring districts to report on the arts instruction provided, student learning in the arts, and providers of arts instruction and by supporting the development of appropriate standards-aligned assessments for use at the state and district levels.
- Rethink instructional time to accommodate the state's goals for meeting proficiency in English language arts and math, while still providing access to a broader curriculum that includes the arts.
- Improve teacher professional development in arts education, especially at the elementary level, and consider credential reforms.
- Provide technical assistance to build districts' capacity to offer comprehensive standards based arts programs.

School and District Leaders

- Establish the infrastructure needed to support arts programs by developing a long-range strategic plan for arts education, dedicating resources and staff, and providing for the ongoing evaluation of arts programs.
- Signal to teachers, parents, and students that the arts are a core subject by providing professional development for teachers and establishing assessment and accountability systems for arts education.

Parents

- Ask about student learning and progress in the arts and participate in school and district efforts to improve and expand arts education.
- Advocate for comprehensive arts education at the state and local levels.

INTRODUCTION

Although California's Education Code calls for all students to be offered a course of study in the four arts disciplines (dance, music, theater, and visual arts), findings from *An Unfinished Canvas* reveal a large gap between policy and practice. California's State Board of Education adopted content standards for the visual and performing arts in 2001, but the Education Code neither requires schools to follow state arts content standards (Section 60605.1b) nor mandates any student assessment in the arts (Section 60605.1c). Recent experience has shown that large-scale assessment used for the purpose of accountability can be effective as a force for implementing standards-based K–12 curriculum and instruction in mathematics, science, social studies, and English/language arts (Pedersen, 2007; Herman, 2007). It is not at all clear, however, whether large-scale assessment could or should be used to support the implementation of K–12 standards-based arts education (Schultz, 2002; Mishook & Kornhaber, 2006). On the one hand, the absence of large-scale arts assessment in California's K–12 accountability system might weaken the status of the visual and performing arts as core subjects in that curriculum. On the other hand, there are few models for large-scale accountability assessment in the arts and only a few states have attempted to implement them.

The development and implementation of effective large-scale assessment in the arts are challenging on many fronts. Some arts educators are concerned that important aspects of achievement and performance in the arts may be subverted by standards and assessments that presuppose that "artistic activity and its products can be deconstructed into discrete components" (Boughton, 2004, p. 589). One-time, on-demand assessments may capture only a small part of what is taught and learned in the arts. Judgment about the quality of student work, especially in the performing arts, may require "real-time" observations and multiple judges. Authentic assessment involving evaluation of extended artistic performances and complex visual and musical products presents unique logistical challenges for large-scale assessment (Myford & Sims-Gunzenhauser, 2004). Even with authentic assessment, efforts to achieve reliability may lead to measurement of things that arts educators do not really care to measure—focusing measures on "the quality of work rather than on the quality of the mind developed through the educational process" (L. Hetland, personal communication, June 15, 2008).

This paper provides a review of the status of large-scale arts assessments and current practice in statewide arts assessment for the purpose of K–12 education accountability. We begin with an overview of the recent history of developments in standards, assessment, and accountability in arts education. Next, we describe and discuss the strengths and limitations of several influential approaches to large-scale arts assessment, the NAEP Arts Assessment and two large-scale portfolio assessment models, the IB arts portfolio assessment and the AP Studio Art portfolio program. Then follows a discussion of the very different approaches to standards-based arts assessment adopted by the five states that currently have active large-scale arts assessment programs: Kentucky, Washington, Rhode Island, South Carolina, and Minnesota.[1] Unless a reference to another source is

[1] California is not among the states currently developing large-scale assessment systems for arts education. Instead, as in many states, efforts are under way to assist schools and districts with developing local standards-based arts assessment. The California Arts Assessment Network (CAAN) was formed by the California Department of Education (CDE) in partnership with The California Arts Project (TCAP) in 1998 and since then has been working with several counties and districts to develop and disseminate arts assessment tools. CAAN provides resources to help California schools and districts develop local arts assessment systems. It provides its affiliate schools and districts with assessment samples, models for training school and district staff on use of the assessments, and a sample item pool. In partnership with TCAP, CAAN has a website, "Student Work Online," where teachers can post student work and obtain assessment feedback from other members. CAAN and similar arts assessment support networks in other states are not

indicated in the text, all comments on the strengths and limitations of the arts assessments reviewed here are the opinions of the authors and are based on our understanding of criteria for technical quality, feasibility, and fairness in large-scale educational assessment. The concluding section of the review examines the NAEP, IB, and AP models of large-scale arts assessment and the five examples of state-level arts assessment through the lens of criteria for a balanced assessment system (as defined by the National Research Council, 2001) as a way of summarizing lessons learned and prospects for an effective arts assessment and accountability system in California and elsewhere. A balanced assessment system is one that combines features of comprehensiveness (covers all important standards and learning goals), coherence (reflects a common understanding of learning in the discipline that links assessment to curriculum and instruction and guides classroom-based as well as district-level or state-level assessments), and continuity (enables monitoring of learning progress over time) (National Research Council, 2001). No existing assessment model or system strikes a perfect balance in terms of these ideal features. Compromise and trade-offs are inevitable. The vision of a balanced approach to large-scale arts assessment—one that captures evidence of important aspects of student learning in the arts and that also supports delivery of strong, sequential, standards-based arts instruction—is not yet a reality .

developing large-scale arts assessment programs and therefore are beyond the scope of this review.

STANDARDS, ASSESSMENT, AND ACCOUNTABILITY IN ARTS EDUCATION: A RECENT HISTORY

In 1994, Congress passed the *Goals 2000: Educate America Act*, marking the culmination of a decade-long push for the development of voluntary national education standards in core subjects. In the same year, the *National Standards for Arts Education* (Consortium of National Arts Education Associations, 1994) and the National Assessment of Educational Progress (NAEP) *1997 Arts Education Assessment Framework* (National Assessment Governing Board, 1994) were published. The *National Standards for Arts Education* were developed to describe what K–12 students should know and be able to do in four arts disciplines—music, visual arts, theater, and dance. Although the content and quality of the standards have been the subject of many debates (see, for example, Ross, 1994; Colwell, 2003), the completion of the standards was a major step forward in building recognition for the arts as part of the core academic curriculum in K–12 education.

Throughout the 1990s, many states developed standards for the visual and performing arts. In 1995, the National Endowment for the Arts (NEA), the U.S. Department of Education, the National Assembly of State Arts Agencies, and the Council of Chief State School Officers (CCSSO) joined with more than 100 national organizations to form the Arts Education Partnership (AEP). The AEP was founded to influence educational policies and promote quality arts education, a role it has continued to play to the present. In keeping with this mission, the AEP maintains the *States Arts Education Policy Database* (see http://www.aep-arts.org/database/index.htm) containing information on the status of state arts education, standards, and assessment requirements. The current database (2006–2007) indicates the existence of some form of state standards for the visual and performing arts in 47 states and the District of Columbia.

The push for development of education standards in the 1980s and 1990s (National Commission on Excellence in Education, 1983; National Education Goals Panel, 1994) was accompanied by calls for new forms of assessment to better match new definitions of student learning and achievement (National Council on Educational Standards and Testing, 1992). In particular, as new curriculum standards incorporating higher order thinking and integrated skills were developed, new methods of performance-based assessment were championed as a way to provide more direct and more visible evidence of learning on the new more complex learning goals than traditional educational testing (National Commission on Testing and Public Policy, 1990; Resnick & Resnick, 1992). The models of performance assessments that were developed to align with new learning standards differed from traditional testing in the definition of constructs to be measured (applied knowledge and integrated skills as opposed to discrete knowledge and skills) as well as in the ways that test items were constructed and scored. In place of multiple-choice, matching, and fill-in-the-blank test items, performance test items were often complex, multistep tasks. Scoring in performance assessment typically entailed use of a multipoint rubric (for descriptions of performance assessment see Herman, Aschbacher, & Winters, 1992; see also Baker, 1998; Wiggins, 1998). While acknowledging the potential of performance assessment methods to capture evidence of learning aligned with new standards, psychometricians and policy analysts also noted the limitations of performance assessment methodology for large-scale assessment of student learning. These limitations include the reliability of scoring procedures and the feasibility and costs of

administering time-consuming performance tasks on a large scale (Linn, Baker, & Dunbar, 1991; Linn, 1993), the potential bias resulting from unfamiliarity with the new test formats (Darling-Hammond, 1994), and a broad range of technical validity and reliability issues concerning test construction, interpretation procedures, and suitability to inform high-stakes decisions (see National Center for Education Statistics, 1996).

The 1997 NAEP Arts Assessment was an early application of performance assessment methods on a large scale (see National Assessment Governing Board, 1994; Pistone, 2002; and discussion of the 1997 NAEP Arts Assessment below). The involvement of state education departments in the development of performance assessment items for the 1997 NAEP Arts Assessment spurred interest in the use of performance assessments for state arts testing programs. In the spring of 1994, the CCSSO was awarded a contract from the Educational Testing Service (ETS) to develop assessment items for the NAEP Arts Assessment. Through this award, 15 state education departments were funded to create prototype performance assessment items in dance, music, theater, and visual arts. In 1995, the CCSSO created the State Collaborative on Assessment and Student Standards Arts Education Consortium (SCASS/Arts) to assist states in developing arts standards and assessments. In 1999, SCASS/Arts published *Arts Assessment: Lessons Learned from Developing Performance Tasks*, a handbook that contained model arts performance exercises in music, dance, theater, and visual arts. These were exercises that had been developed for the 1997 NAEP Arts Assessment but had not been selected for use. The items were aligned with the National Arts Standards and were revised by SCASS/Arts members to better reflect grade-level expectations for grades 4, 8, and 12.

Several states that had participated in the NAEP Arts Assessment development moved quickly to develop their own statewide arts assessments. At the time that Yan and Rieder (October, 2001) completed their review of the status of large-scale arts assessment in the states, nine states had mandated assessments in the arts. At that time, five states had implemented required arts assessments (Kentucky in 1993, Illinois in 1997, and Maine, Minnesota, and Oklahoma in 1999), three more states had plans for implementing state arts assessments in the near future (Missouri in 2002, New York in 2002–2003, and Washington in 2008), and New Jersey planned to do so at a time to be determined. One more state, Maryland, had commissioned the 2001 review of state arts assessment as part of its own planning for a future state arts assessment. Yan and Rieder also identified a number of states (including Alaska, California, Massachusetts, Ohio, Pennsylvania, Utah, and Vermont) that did not have mandates for state arts assessments but were "encouraging activities in fine arts assessment at the local level."

The high point of the movement toward state arts assessments at the turn of the millennium may have come in 2002. At that time, Kentucky (grades 5, 8, and 11), Missouri (grade 5), and Maine (grades 4, 8, and 11) each had a stand-alone statewide arts assessment consisting of a combination of multiple-choice and constructed-response test items covering all four arts subjects (dance, theater, music, and visual arts). Oklahoma had a stand-alone statewide multiple-choice test covering visual arts and music for students in grades 5 and 8, and Illinois administered multiple-choice items covering all four arts subjects as part of its statewide social sciences test in grades 4 and 7 and as a stand-alone arts assessment in grades 9/10. Also in 2002, drawing on the example of the 1997 NAEP Arts Assessment and the work of the SCASS/Arts Education Consortium, the AEP and the CCSSO published *Envisioning Arts Assessment: A Process Guide for Assessing Arts Education in School Districts and States* (Pistone, 2002). The author, Nancy Pistone, noted that at the time she was writing, "17 states have initiatives in place that support statewide arts assessment and still others are pending" (2002; p. 11).

However, even before the publication of *Envisioning Arts Assessment* in September 2002, the tide had turned. With the signing into law of *No Child Left Behind* (NCLB) in January 2002, momentum in developing statewide arts assessments was all but halted. Although NCLB confirmed

the status of visual and performing arts as core curricular content, it also introduced new requirements for accountability and assessment that were not favorable to statewide arts assessment. With states facing unprecedented pressure to demonstrate adequate yearly educational progress in other core subject areas, particularly mathematics and reading/English language arts, public attention and school resources were increasingly deflected away from the arts. The need to direct relatively more time and resources to high-stakes assessments in other core subject areas coupled with a clear understanding of the level of resources needed to create high-quality large-scale arts assessments caused many states to put plans to develop assessments in the arts on the back burner. Illinois, for example, which had been among the first states to adopt state standards for the arts and had initiated a statewide arts assessment, dropped its statewide arts assessment in 2004 for budgetary reasons (Illinois Creates, n.d.).

Today, very few states are pursuing large-scale efforts to assess arts learning. Indeed, the test-based accountability requirements of NCLB have resulted in reduced interest in use of performance assessment and expansion of multiple-choice elements in state testing programs (GAO, 2003; Darling-Hammond & Rustique-Forrester, 2005). As a result, one effect of NCLB has been a resurgence of long-held concerns about the negative impact of "too much testing of the wrong kind" in schools (Barton, 1999). While advocates of standards-based reforms in the 1990s called for the development of "tests worth teaching to" (Resnick & Resnick, 1992) as a means of improving the quality of instruction and instructional outcomes, post-NCLB critiques of test-based accountability have pointed to negative effects of testing that include narrowing of the curriculum (teaching to the test and not to the standards), loss of instructional time to test preparation, less instruction on complex reasoning and performance (Haertel & Herman, 2005; Darling-Hammond & Rustique-Forrester, 2005), and decreased instruction in untested subjects such as the arts (Herman, 2004; Bodilly, Augustine, & Zakaras, 2008).

At present, Kentucky is the only state with a mandated statewide arts assessment. Washington has piloted and will soon implement a statewide classroom-based performance assessment for arts education. Rhode Island is defining visual and performing arts proficiencies and a variety of arts assessment methods as part of new proficiency-based graduation requirements for its high schools. South Carolina has developed an innovative web-based approach to large-scale arts assessment that is currently used only in schools receiving state grants for arts education. Minnesota is following the lead of Queensland, Australia, in developing a school-based approach to arts assessment. The very different approaches to large-scale arts assessment and accountability taken by each of these five states are described in a later section of this review.[2] The next section provides an overview of three long-established arts assessments that exemplify important aspects of good design for large-scale performance-based assessment in the arts.

[2] In addition to the five state-level arts assessments described in this review, Maryland currently maintains a commitment to implementing a state-mandated assessment of quality in arts education (known as *Portfolio Plus*, see the *Introduction to Maryland Fine Arts Education* at http://mfaa.msde.state.md.us/source/MDFAeducation_1e.asp). Arts assessments in Maryland will be aligned with state content and achievement standards. However, the final form that the Maryland arts assessments will take was not known at the time of the writing of this review.

LARGE-SCALE ARTS ASSESSMENT

Designing effective large-scale performance assessments poses significant technical challenges for any subject area. For the visual and performing arts, however, the desirability of a performance-based approach over traditional educational testing, which measures discrete knowledge and skills, is readily apparent. The evaluation of artistic learning and proficiency, whether in theater, dance, music, or the visual arts, has traditionally entailed expert judgment of performance and performance products (e.g., music recitals and visual art exhibitions). Indeed, one of the most prevalent performance assessment methods, portfolio assessment, was inspired by evaluative practices that originated in the visual arts. Two of the three examples of large-scale performance-based arts assessment systems described here make use of the portfolio assessment method: the International Baccalaureate (IB) portfolio assessment and the Advanced Placement (AP) Studio Art portfolio. We begin with the third important example of large-scale performance assessment in the arts, the 1997 National Assessment of Education Progress (NAEP) Arts Assessment.

THE NAEP ARTS ASSESSMENT

Instituted in 1969, the NAEP performs ongoing assessments of what U.S. fourth-, eighth-, and twelfth-grade students know and can do in key subject areas such as reading, mathematics, writing, U.S. history, science, the arts, civics, and geography. The NAEP does not produce scores for individual students. Instead, a matrix sampling approach is used to randomly select representative samples of students to whom small subsets of items from an overall test are administered. Statistical techniques are then used to combine results and to provide data on the achievement of selected national subpopulations of students. The result is a "national report card" that provides comparable data on student achievement across states that are not obtainable from any other source. (Background information on the NAEP in this section was obtained from the NAEP website at http://nces.ed.gov/nationsreportcard/.)

In its history, NAEP has conducted a series of arts assessments: for music in 1972, for visual arts in 1975, and for both in 1978 and 1997. Theater was added to the 1997 NAEP Arts Assessment. A dance assessment was also created in 1997 but was dropped when NAEP could not identify an appropriate national sample of students. An updated version of the 1997 NAEP Arts Assessment is being repeated in 2008. Although the 1970s versions of the test were considered innovative for their time, many of the test items were multiple choice. Like other assessments of the period, the arts assessments focused primarily on assessing arts-related behavioral objectives and content knowledge, although some performance tasks in art design and drawing were included. Foreshadowing challenges test developers face today, NAEP was able to score only about half the performance-based items in the 1978 test because of the high costs of employing judges (Persky, 2004; for a detailed account of the evolution of the NAEP Arts Assessment, see Myford & Sims-Gunzenhauser, 2004).

As is true of any well-designed large-scale assessment, all versions of the NAEP Arts Assessment have been grounded in a fully developed assessment framework that describes in detail the skills and content knowledge to be assessed. The 1997 NAEP Arts Assessment framework (National Assessment Governing Board, 1994) and the administration, scoring, and reporting procedures for the 1997 NAEP Arts Assessment were considerably different from the earlier versions (Persky, Sandene, & Askew, 1998). The newly developed National Arts Standards provided, for the first time, content standards on which to ground the assessment framework. In addition, the assessment provided an opportunity to try out new psychometric tools and techniques for administering and

scoring performance-based tests that had been developed in the intervening years since the 1970s. A large NAEP arts committee made up of psychometricians and arts specialists then collaborated in an extensive 2-year process to create assessment questions and scoring criteria according to the framework specifications (Persky, 2004).

In the 1997 and 2008 versions of the NAEP Arts Assessment, items are organized in blocks of activities centered around one theme or problem and allow students to apply their knowledge and skills simultaneously. Within each block, students are assessed on art processes and content by *creating* or generating original art, *performing or interpreting* an existing work of art, and *responding* to art through observations, descriptions, analysis, or evaluation based on knowledge and skills they had developed in a specific arts discipline. Some items within the block require complex performance tasks. Others require students to write open-ended responses, and some require answering multiple-choice questions. Students take the NAEP Arts Assessments in just one discipline. The time spent on completing the assessment varies. In the music assessment, for example, the general student population completes three blocks of activities and a subsample of students who are currently enrolled in a music activity complete two additional blocks. Exhibit 1 describes the kinds of exercises students complete. Students spend an average of 90 minutes engaged in the NAEP Arts Assessment tasks. Scoring of performing arts tasks is done on the spot by test administrators who are specialists in the arts and are trained in how to observe and score performances. For the visual arts, judges are trained in how to photograph artworks from different angles and the photos are sent to a central location to be scored by trained judges.

Exhibit 1
Examples of Assessment Exercises: 1997 NAEP Arts Assessment Framework

	Kinds of Exercises	
	Creating and/or performing: assessed with performance tasks	Responding: assessed with written exercises and multiple-choice questions
Music	Create and perform a rock-and-roll improvisation on a MIDI keyboard.	Listen to pieces of music and then analyze, interpret, critique, and place the pieces in historical context.
Theatre	Work in a group to create and perform an improvisation about a camping trip.	Listen to a radio play and then do a series of written exercises about staging the play for young children.
Visual Arts	Using markers and a cardboard box, create a package designed to hold a whisper or a scream.	Study artworks and then do exercises exploring aesthetic properties and expressive aspects of the works.
Dance	Work with a partner to create and perform a dance based on the idea of metamorphosis.	Watch ethnic folk dances on videotape and then analyze and place the dances in historical context.

Source: http://nces.ed.gov/nationsreportcard/arts/whatmeasure.asp

Strengths

The 1997 NAEP Arts Assessment was heralded by many as an unprecedented achievement in performance-based arts assessment (see Persky, 2004; Schneider, 2003). Developed by the ETS, it incorporated state-of-the-art measurement, administration, scoring, and reporting techniques and procedures to ensure the high validity and reliability of its test results. The arts assessment framework represents the combined efforts of skilled NAEP psychometricians as well as the expertise of national leaders in arts education, including many developers of the National Arts Standards. It successfully carried out many ambitious goals, including building a performance assessment based on arts content standards; using complex, applied performance-based tasks to

recognize and measure creative achievement in the arts; and adhering to strict administration guidelines and scoring criteria. The NAEP Arts Assessment has served as a model for many state arts assessment initiatives. Arts assessment specialists continue to draw upon the NAEP framework for arts item development, assessment administration, and scoring. As noted, in 2002 the CCSSO and the AEP collaborated to produce *Envisioning Arts Assessment* (Pistone, 2002), a guidebook outlining the steps for developing a standards-based large-scale performance assessment in the arts based on the framework for, and lessons learned from, the NAEP assessment.

Limitations

Despite its successes, many criticisms have been made of the 1997 NAEP Arts Assessment (see Bonbright & McGreevy-Nichols, 1999; Colwell, 1999; Donmoyer, 1999; Eisner, 1999). The inclusion of creating or performing exercises (performance assessments) in the 1997 NAEP Arts Assessment significantly decreased the numbers of students in the assessment because of the increased time and expense to develop, administer, and score the items. The 1997 NAEP visual arts and music assessments were administered to nationally representative samples of eighth-graders in public and private schools and were completed by 2,999 (visual arts) and 2,275 (music) students, respectively (Persky et al., 1998). The NAEP theater assessment was administered to 1,386 students, in this case limited to students who had completed at least 30 hours of classroom instruction in theater by the end of the 1996–97 school year (Persky et al., 1998). A similar requirement of classroom instruction in dance caused the elimination of the dance assessment items in 1997 because too few eighth-graders students were enrolled in dance classes (Persky et al., 1998). The fact that performance assessments take more time to administer also reduced the total number of items that could be included and the range of content that could be covered in the 1997 NAEP Arts Assessment. The limited number of items (and content covered) could be seen as a threat to the test's validity given the broad range and variations in the content of arts education in U.S. schools (Eisner, 1999). Moreover, the entire assessment process was very expensive—in fact, the most expensive NAEP assessment ever conducted (Morton, 1999).

Many questions also were raised about whether states could replicate an assessment process similar to the 1997 NAEP Arts Assessment for the purpose of accountability. Because the primary purpose of the NAEP is to provide a national profile of achievement levels and skills rather than to test students individually, it is possible to use a matrix sampling approach. Only a subsample of students take the test and, within this subsample students are required to complete an assessment in only one of the four arts areas. Although matrix sampling as in the NAEP may be seen as strength given that it reduces the time and expense of administering performance assessment items on a large scale, it also presents challenges in monitoring individual student or school results for accountability purposes (Schneider, 2003).

THE INTERNATIONAL BACCALAUREATE (IB) ARTS PORTFOLIO ASSESSMENT

The IB program of study is in use in over 60 countries, with many programs in the United States. The IB arts portfolio assessment is just one of a large number of subject area assessments developed by this organization. These highly regarded assessments are accepted by major universities throughout the world, including a large number of institutions of higher education within the United States. In addition to the studio arts portfolio assessment, IB provides curricula and assessments in the performing arts, but the performance portion is conducted at individual schools rather than through the external examinations used for studio art portfolios. (All information on the International Baccalaureate arts assessments comes from the IB website at www.ibo.org.)

Although the IB program provides curricular guidelines, schools are encouraged develop their own program of instruction. The curricula and assessments are international and not strictly based on

Western culture. Because of the many countries in which they exist, schools are given wide flexibility to accommodate varied cultural contexts. A central Curriculum Office evaluates a school's IB program to ensure that it is meeting the criteria.

All students in the visual arts at the Diploma Programme level (ages 16–19) must complete a studio and research portfolio for assessment. In the IB program, the portfolio assessment is used to determine individual student achievement and to graduate the student. The results of the arts portfolio assessment are not reported to the IB program for purposes of evaluating the school.

Students are encouraged to add to their portfolios throughout their course of study. Portfolios include not just the students' best finished works, but also working pieces demonstrating the process of their visual research, including work in progress, sketches, and a research notebook with critical self-reflections. An IB examiner, typically an experienced IB visual arts teacher, evaluates the portfolio for the following five criteria: imaginative expression, purposeful exploration, meaning and function, formal qualities, and technical and media skill. As part of the assessment process, the examiner interviews the student.

After the examiner's visit to the school, photos of the artwork and photocopies of other materials in the portfolio are sent to a central location in Cardiff, Wales. There, a group of trained moderators compare the visiting examiner's judgments against agreed-upon performance benchmarks. Performance benchmarks are written performance descriptors coupled with actual samples of student work that moderators have identified as exemplifying designated achievement levels. The student work samples used as performance benchmarks are drawn from the work of students from around the world. They are posted each year on a virtual gallery that all teachers and students in the program can access on the World Wide Web.

Strengths

In his favorable review of the IB approach to arts assessment, Doug Boughton (2004) argues that collecting evidence of student performance in a portfolio is a much more authentic and appropriate measure of arts learning than paper-and-pencil tests, which tend to artificially dissect the artistic process. The contents of the portfolio used in portfolio assessment are products of ongoing instructional activities. Developing a portfolio and keeping a research notebook encourages students to become more critical and self-reflective. The fact that students select their own entries for the portfolio and notebook promotes their artistic independence, autonomy, and exploration. Many teachers contend the process also increases students' motivation. In addition, the portfolios can provide the teacher with a unique window into individual students' learning (for additional details, see Boughton, 2004).

The processes of *benchmarking* and *moderation* used in the IB assessment model are additional strengths. Using sample student artwork as benchmarks ensures validity and fairness by providing examiners and students with a common understanding of what student work looks like at different achievement levels (Boughton, 2004). This has evolved into a more transparent assessment process where common agreements about what exemplary work and the qualities it embodies can be shared throughout the educational system. Moderation—external review of assessment scoring by specially-trained examiners—not only ensures that judgments of student work are more reliable and fair, but also develops agreement on standards among the examiners and teachers and indirectly provides teachers as well as students with feedback. Becoming trained as an examiner provides an opportunity for experienced teachers to advance their understanding of the field.

In addition to ensuring valid and reliable assessments of student artwork, moderation and benchmarking also encourage debate about student work, which, according to a former examiner for the IB portfolio assessment, "is essential in an assessment context where students are required to push the limits of their own understanding, to take risks, exercise imagination, and interpret the

visual world critically" (Boughton, 2004). The strength of the portfolio assessment in this context is that it evaluates the student's capacity to go beyond the standards and use the higher order thinking skills that are inherent to the arts. Teachers must be cognizant of the portfolio criteria when they develop curricula and instructional plans, because their students need to complete the portfolio satisfactorily in order to graduate with an IB diploma. Therefore, the assessment criteria are a driving force in curricula and instruction.

Limitations

The IB arts portfolio assessment has several limitations in common with the 1997 NAEP Arts Assessment. Like the NAEP assessment, this portfolio assessment process is costly to implement. A trained examiner must go to the school and to ensure reliability the artwork is sent to a central location where additional examiners must be employed to judge the student work. This process requires both time and expense that are not incurred with pencil-and-paper assessments. Also as in the NAEP Arts Assessment, not all students are assessed. In the IB arts portfolio assessment, only students at the senior level are evaluated. If implemented on a larger scale for accountability purposes, the process of locating and training experienced examiners, coordinating the on-site interviews, and handling the logistics of transferring student work to a central location for further moderation would be considerably more challenging.

THE ADVANCED PLACEMENT (AP) STUDIO ART PORTFOLIOS

The Advanced Placement (AP) program of the College Board recently celebrated its 50th anniversary in U.S. schools. AP examinations are currently given in 35 subject areas. As is true of all the AP exams, the key mission of the AP Studio Art portfolios is to demonstrate that the high school student has met the equivalent standards of college-level learning in a given subject area. AP Studio Art program offers three portfolios: drawing, 2-D design, and 3-D design. In 2005, a total of 24,254 students took the AP Studio Art portfolios; the majority were 12th-graders (College Board, 2006). Although they have changed over the years, the AP Studio Art portfolios have been given every year since 1972. The AP Studio Art portfolios are the only AP exams that are not a written test. Art history and music theory are among the subjects with written AP exams. (Background information on the AP Studio Art program in this section was obtained from the College Board website at http://www.collegeboard.com/student/testing/ap/sub_studioart.html.)

Rather than receiving standards or specific curriculum guidelines, AP Studio Art teachers are provided with portfolio guidelines and expected to develop a course that will enable students to complete the requirements. The *AP Studio Art Course Description* (College Board, 2007a) is given to teachers each year. In addition, the College Board produces a color poster that includes exemplary artwork from the previous year and an abbreviated version of the portfolio guidelines. The AP portfolio guidelines require that the course focus on three main areas in art making regardless of which portfolio the student chooses: quality, concentration (for which students must submit 12 slides that document an investigation of a visual idea or problem and a short written statement about the work), and breadth (for which 8–16 slides must be submitted that demonstrate a wide range of experiences in the concentration). Students also submit five works that they feel best represent their artwork. In May of each year students are required to submit a portfolio, which consists of at least a year's work. Students may submit work completed out of school in addition to work completed in their high school class.

Unlike the IB portfolio assessment, the AP exam requires that the student portfolio meet very specific criteria. Each area (quality, concentration, and breadth) is scored according to these criteria on a scale of 6 (excellent) to 1 (poor). Teachers receive a scoring guideline so they understand the criteria to be used to evaluate student work and what quality of work is expected at each scoring

level. The criteria are qualitative. The following are excerpts from the *AP Studio Art 2007 Scoring Guidelines* (College Board, 2007b) for 2-D design quality:

> **6 Excellent** Quality. Work at this level … shows an imaginative, inventive, and confident articulation of the principles of design …
>
> **5 Strong** Quality. Work at this level … demonstrates a strong grasp of the elements and principles of design, using them to express a visual idea …
>
> **4 Good** Quality. Work at this level … demonstrates a good understanding of the elements and principles of design…
>
> **3 Moderate** Quality. Work at this level … shows an emerging understanding of the elements and principals of design…
>
> **2 Weak** Quality. Work at this level … shows little understanding of elements and principles of design …
>
> **1 Poor** Quality. Work at this level … shows no apparent understanding of the principles of design …

The requirements for implementation are similar to those of the IB portfolio assessment. Experienced art teachers and college art faculty serve as portfolio evaluators called readers. During the evaluation process a team of a chief reader, exam leaders, and table leaders evaluate the portfolios. The teachers selected as readers must have at least 3 years of experience as a studio art teacher or be a faculty member who teaches introductory-level college studio art. Before the evaluation process begins, readers receive training in evaluating sample works. During the evaluation, if the scores are more than 3 points apart in a given section of the portfolio that section is sent to the team leaders for resolution. To meet the statistical demand for high score reliability, only a single AP grade is reported for studio art rather than reporting subscores. Scoring guides for the portfolios can be found on the AP website (http://www.collegeboard.com/prod_downloads/ap/students/studioart/ap07_sg_studioart.pdf).

Strengths

The strengths of the AP Studio Art portfolios are similar to those of the IB portfolio assessment: Portfolio assessment captures aspects of student performance that are critical to artistic development, the individual artist's creativity is not compromised as it is in more standardized forms of assessment, the assessment process promotes the learning experience, the process of moderation ensures fairness and reliability, and benchmarking provides a transparent and effective means to make clear what exemplary work is in the studio arts. The AP Studio Art portfolios have been demonstrated to be successfully implemented with thousands of students over many years and to work within the typical U.S. public high school.

Limitations

Even more so than the IB arts portfolio assessment, curriculum and instruction in the AP Studio Art portfolios are driven by the portfolio criteria. Students must submit a portfolio that meets very specific requirements, and therefore the course must be dedicated to preparing work for the portfolio. Although the AP program strives to avoid teaching to the test, teachers may have a natural tendency to narrowly focus teaching approaches and activities to support students' production of works that will receive high scores rather than to emphasize students' acquisition of the knowledge, skills, and abilities needed to independently produce high-quality work. According to Lois Hetland (personal communication, June 15, 2008), "The AP privileges the elements and principles of design over other aspects of artistic development in ways that often diminish learning of those other vital areas."

Like the IB portfolio assessment, the AP Studio Arts exam is not used for the purpose of school accountability. It also is given to students only once in their high school years, most typically at the senior level. Implementing the exam more widely for the purpose of accountability would be very expensive. Large numbers of experienced reviewers who could make reliable judgments of artworks would need to be identified and trained. Bringing together teams of reviewers could prove to be logistically complex. Since portfolios are produced over time with materials available at the school, if assessments are to be valid all schools would need to have equal access to the materials needed to produce work for the portfolio. Additionally, student work would need to be mailed and stored. Although technology makes it possible to eliminate some of this expense, since portfolios can be digitally photographed and sent through the internet or on CDs (Boughton, 2004), all schools would need the technology and know-how to produce and transmit digital portfolios.

STATE ARTS ASSESSMENT SYSTEMS

As noted in our review of recent history in standards, assessment, and accountability in arts education, the high point in statewide arts education assessment may have been reached in 2002. In the years following the passage of NCLB, the number of states mandating statewide arts assessment rapidly declined. (Information on the status of state arts assessment policy was obtained from the Arts Education Partnership's online *States Arts Education Policy Database* at http://www.aep-arts.org/database /index.htm.)

- Beginning in 1998, Illinois administered arts assessment items as part of the statewide social studies test in grades 4 and 7 and as a stand-alone test in grades 9/10, aligned its assessment with state arts standards in 2001, and then abandoned statewide arts assessment in 2004 because of budget cuts.
- In 1999, Oklahoma administered its Arts Criterion Referenced Test to all students in grades 5, 8, and 11; reduced the administration to grades 5 and 8 in 2000; and then in 2003 eliminated the statewide arts test and required school districts to develop local art and music assessments.
- In 2002, Missouri initiated a statewide arts assessment but then suspended it in 2004 because of budget cuts.
- At present (2008), Maryland, Maine, New Jersey, and New York[3] have delayed or suspended earlier efforts to implement statewide arts assessments.

The only functioning model of state-mandated, large-scale arts assessment to be found currently in the United States is in Kentucky. The Kentucky system is an example of traditional (as opposed to performance-based) assessment, and it is the first of the large-scale state approaches to arts assessment described here. The second large-scale state arts assessment described is Washington's Classroom-based Performance Assessment (CBPA) system, a performance assessment approach. The CBPA system will include an arts component and will be mandated for all schools in Washington beginning in the 2008–09 school year. The third noteworthy state arts assessment system is the new Rhode Island high school diploma system that requires students to demonstrate proficiency in six core areas, including the arts. Beginning with the graduating class of 2008, new proficiency-based graduation requirements (PBGRs) have been put into effect. The PBGRs present schools and students with multiple options for demonstrating proficiencies in core academic areas and in applied learning skills. The last two examples of state arts assessment systems are not mandatory for all schools and students but are nonetheless ambitious and innovative examples of large-scale arts assessment. South Carolina has developed and implemented a web-based system that combines multiple-choice and performance assessments aligned with state standards for the visual and performing arts. Minnesota is in the process of adapting a school-based arts assessment and accountability system developed in Queensland, Australia.

[3] New York put its arts assessment program on hold because of resource allocation issues, although it is finalizing plans to implement voluntary ninth-grade arts assessments in 2008–09 (E. Marschilok, personal communication, April 17, 2008).

KENTUCKY

Since the 1990s, Kentucky has had an accountability system that measures student progress in the seven core content areas, one of which is arts and humanities. The state mandates that students be instructed and assessed in *all* core content areas, so it developed a large-scale assessment in the arts. The arts assessment is one component of the Kentucky Core Content Test (KCCT). The purposes of the KCCT are to measure how students and schools are performing, determine where students need help, and inform instructional decision-making. Each of the core content areas makes up a different percentage of the test. The arts and humanities assessment comprises just a small portion of the KCCT and is administered only in grades 5, 8, and 11. At the elementary level, the arts constitute 5% of the assessment, while mathematics and reading are 22% each and social studies and science are 14.5% each. At the middle school level, the arts make up 6.75% of the assessment, whereas reading and mathematics each make up 19% and social studies and science are 15% each. At the high school level, the arts consist of 7% of the assessment, and reading, mathematics, social studies, and science are 14% each (Kentucky Department of Education, 2007) (Background information on the KCCT in this section was obtained from the Kentucky Department of Education website http:kde.state.ky.us/KDE/Instructional+Resources/Curriculum+Documents+and+Resources/Core+Content+for+Assessment/).

The arts portion of the KCCT consists of eight multiple-choice questions and two open-response items, which can be in any of the four art disciplines (music, visual art, theater, and dance). Exhibit 2 shows sample items from the arts portion of the KCCT. The questions are designed to cover the Core Content for Assessment, which is the equivalent of state content standards. In Kentucky, the core content in the arts is divided into five subdomains: structures in the arts, humanity in the arts, purposes for creating the arts, processes in the arts, and interrelationships among the arts. Of these five areas, three are assessed in statewide tests. Interrelationships among the arts is taught only at the high school level and is considered a support for the other standards. Processes in the arts cannot be assessed through written response because it involves creating and performing art (Kentucky Department of Education, 2006). The content subdomains included in the state assessments are those that ask students to describe, analyze, evaluate, and explain the arts. The makeup of the state arts assessments is 45–55% in the structures subdomain, 10–15% in the purposes subdomain, and 30–45% in the humanities.

Exhibit 2
Sample Items from KCCT Arts and Humanities

Fifth grade, open response:
The Dancing Animals
7. Your class is going to make up a play about animals that can dance. Your job is to create two characters for the play.
   a. Name TWO characters that could be in a play about dancing animals. Describe what the characters would look like and what they would wear.
   b. For EACH of the two characters, describe the character's part (what the character does) in the play.

Eighth grade, multiple choice:
3. In theater, if an audience has the same feelings or reactions as those of a character, the audience is experiencing
   a. empathy.
   b. motivation.
   c. suspense.
   d. spectacle.

Eleventh grade, multiple choice:
4. Complementary colors are two colors at opposite points on the color wheel. Which colors are complementary?
   a. red and orange
   b. green and yellow
   c. violet and green
   d. orange and blue

Source: Kentucky Department of Education Sample Release Questions (Kentucky Department of Education, 2006).

Strengths

The fact that the KCCT is a traditional paper-and-pencil test is very advantageous from a cost perspective. Although content area expertise is required for developing arts assessment items, the costs of developing the multiple-choice and simple open-ended response item formats are relatively low compared with those of developing performance assessment items. Likewise, scoring and reporting of arts items on the KCCT can be handled efficiently and in the same manner as items from other areas of the curriculum. Scoring of open-response questions requires training expert raters in the use of a scoring rubric that evaluates how complete and thorough the student's response was, but here again costs are low relative to the level of training required for scorers of more complex performance tasks.

Inclusion of arts items in the KCCT is a strength in that it puts the arts on par with other subjects that are assessed. Just as with other areas of the curriculum, a school's performance on the arts component of the state assessment is used to determine whether the school is eligible for rewards, sanctions, or improvement measures. For example, if a school performs below standard in the arts, it is subjected to teacher evaluations and entitled to improvement assistance from the state in the form of consultation from an arts expert (Kentucky Department of Education, n.d.).

Limitations

The impact of the KCCT arts assessment on the delivery of standards-based arts curriculum has been mixed. On the positive side, more schools are including the arts as a requirement in their curriculum and teachers are aware of the requirement to provide arts instruction. Most classroom teachers, however, leave arts instruction to a specialist or to the fifth-grade teacher who prepares students for the exam. Some districts have gone to a humanities model, in which one "arts and humanities" specialist is hired to cover all four arts disciplines and prepare students for the exam (personal communications, D. Horn, March 9, 2007; P. Shepherd, March 15, 2007; and

D. Thurmond, March 21, 2007). While more arts instruction is being provided, some Kentucky educators are looking to redevelop the statewide arts assessments to address the performance and creativity components of the standards (personal communication, D. Horn, March 9, 2007). Furthermore, a survey of teachers in Kentucky found that they believe they will continue to place more emphasis on subjects that are more heavily weighted in the state assessments. Because the arts count for only a small portion of the overall score, many elementary teachers reported that they are less inclined to give time to the arts in their curriculum (Horn, 2005).

WASHINGTON

Washington has a standards framework similar to most states', which includes the arts among the core academic disciplines. Instruction in all core academic disciplines is mandated by state law, and in 2001 a task force was created to support schools in meeting the requirement to provide arts instruction. The push for arts assessments started in 2003, when the Arts Assessment Leadership Team was assembled to develop Classroom-based Performance Assessments (CBPAs) in the arts. There are currently CBPAs in the arts, social studies, and health/fitness—the areas of the curriculum that are not covered in Washington Assessment of Student Learning (WASL). The WASL is the central component of the state assessment system and measures student learning in mathematics, language arts, and science. Unlike the WASL, the CBPAs are designed for the purpose of supporting the delivery of standards-based arts instruction, rather than as measures of student learning outcomes in the arts. (Background information on arts assessment in Washington in this section was obtained from the Washington Office of the Superintendent of Public Instruction website at http://www.k12.wa.us/Assessment/WASL/Arts/default.aspx.)

In 2004, a state law was passed to put some substance behind the mandate to provide arts instruction. Specifically, it requires that students complete coursework in the arts in order to graduate from high school and that the content of the course meet the highest benchmark (Benchmark Three) in the state learning standards. Requiring high school students to have arts instruction at the Benchmark Three (graduation requirement) level entails a further obligation to provide all Washington public school students with earlier arts instruction and learning opportunities at the Benchmark One (grade 5) and Benchmark Two (grade 8) levels (Washington State Arts Commission, 2006). The law also stipulates that all schools implement CBPAs in the arts starting in 2008–09 (assessments are designed for grades 5, 8, and high school graduation, and schools need to do assessments only in one art form). Schools will be required to use the CBPA to demonstrate that their students are meeting standards in at least one art form and to submit an implementation verification form to the state Office of Superintendent of Public Instruction (OSPI).

Washington's classroom-based assessments are actual arts activities integrated into the curriculum. Students' tasks in the assessment are to create and perform and then respond by reflecting on their own art-making process. Exhibit 3 shows an example of a CBPA activity script to be used by a fifth-grade teacher. Because the assessment is also an instructional activity, students take class time to develop their piece before presenting it for assessment. The teacher is provided with an instruction and assessment guide to implement the assessment. It includes a script that explains the project that students are to complete and the standards-based criteria by which their work will be evaluated. Teacher guides for the arts CBPA are available on the OSPI website.

Exhibit 3
Example of a Fifth-Grade Visual Arts CBPA Activity Script

You've Got It Covered

A recording company is looking for artwork for the front cover of a new CD of music from around the world. You are a graphic artist. A graphic artist is a person who creates artwork for the purpose of selling a product. Your job is to create a cover design that reflects the music on the CD.

The recording company will provide a piece of music to inspire your artwork. The first time the music is played, you will only listen. As you listen to the music again, you will draw your CD cover with a pencil. Then you will use color to refine the cover for the recording company. The recording company will only use your cover if you include a variety of line types and qualities, shapes, and colors. Following this, the recording company asks you to respond about the choices you made while creating your CD cover.

The recording company explains that you must meet these requirements as you create your cover:

- Create a drawing that is inspired by the music that will be on the CD.
- Use three distinctly different types of lines in your drawing that can be seen clearly (horizontal, vertical, diagonal, curved, and zigzag).
- Use three distinctly different qualities of lines in your drawing that can be seen clearly (thick, thin, broken, smooth, fuzzy, long, short, light, dark).
- Use three distinctly different shapes (geometric shapes and/or free-form/organic shapes).
- Use a variety of colors in your composition with the materials you are provided.
- Use lines, shapes, and colors to organize your composition using symmetrical, asymmetrical, or radial balance.

You will have 30 minutes to finish your CD cover as you listen to the music. The recording company has given you 15 minutes to respond about your CD cover. Be sure to use visual arts vocabulary correctly in your written response.

Source: The Washington CBPA: The Arts, Grade 5, Visual Arts, You've Got It Covered! Full training set. Available from the Washington OSPI website at http://www.k12.wa.us/Assessment/WASL/Arts/default.aspx

Strengths

Several advantages of the CBPA resulted from the work put in to the development phase. First, the assessments require very little arts knowledge or capacity to make qualitative artistic judgments. The guides are thorough in defining art terms and describing what student work should look or sound like. The scoring guide includes rubrics that align with the standards-based criteria of the activity. The criteria are simply counted, so a score of 4 is given to a student who completes four of the criteria, a score of 3 to a student who completes three of the criteria, and so on. Qualitative judgment in this scoring is minimal, so reliability can be attained even when the teacher has limited capacity to evaluate artistic quality. The scoring guide also includes an "anchor set" of student samples that are scored and a "practice set" that the teacher can use to practice scoring.

Another strength of the Washington CBPA is that the assessments take very little time away from classroom instruction. Because they are performance activities that are meant to be integrated into the curriculum, they can be implemented as a normal arts lesson. Also, because teachers assess their own students' work, the CBPA can be implemented on a large scale without creating a central panel to assess the work of all students in the state.

Limitations

The development of the CBPA was time intensive. Items were developed by a panel of arts educators and then piloted through multiple iterations over 5 years to develop two sets of

assessment items (total of 60 items). This panel also sorted and scored several thousand student samples. (See Joseph, 2006, for a complete description of the development of the CBPA.) Another limitation comes from the fact that the CBPA places responsibility for quality control at the teacher and district levels. Each teacher is responsible for administering the arts CBPA to their students and for scoring. The district is then responsible for reporting arts assessment results to the state. The absence of any mechanism to monitor whether or not teachers are accurately implementing and scoring the assessments limits the reliability of assessment results for accountability purposes. Additionally, while teachers benefit from guides and scripts that minimize their need to receive training in the arts, this format limits teacher development in understanding how to teach and assess the arts.

RHODE ISLAND

Beginning with the class of 2008, high school students in Rhode Island must demonstrate proficiency in the arts and in five other core academic areas (mathematics, English language arts [ELA], social studies, science, and technology) as part of a new Rhode Island Diploma System that includes performance-based graduation requirements (PBGRs). Under the new system, districts must develop multiple opportunities for students to demonstrate proficiency in the arts and other core academic areas. In order to graduate, each student must demonstrate proficiency in at least one art form (music, visual arts, theater, or dance). This demonstration of proficiency in the arts can be accomplished through a combination of coursework, "local assessments" (e.g., classroom assignments, on-demand and extended tasks, projects, end-of-course assessments, and other local assessments, of which 50% must be performance based), and "schoolwide diploma assessments" (additional measures including exhibitions, graduation portfolios, and Certificates of Initial Mastery [CIM]). (Background information on the Rhode Island Diploma System was obtained from the Rhode Island Department of Elementary and Secondary Education website at http://www.ride.ri.gov/HighSchoolReform/default.aspx.)

The Rhode Island Arts Learning Network (see http://www.riartslearning.net) is supporting four statewide teams (one for each art form) to assist local districts in developing definitions of proficiencies in the arts. State guidelines call for these definitions to conform to state and national standards and to be realistic and achievable by "all kids" (Rhode Island Department of Education, 2004, p. 25).

Demonstrations of proficiency in the arts will be accomplished by local districts and schools as part of their local proficiency-based assessment systems that are designed to complement and reinforce the state assessment system. Local proficiency-based assessment systems will include specialized assessments that are designed to measure deep content knowledge and "habits of thinking" within the context of extended work by individual students. Specialized assessments include the following:

- Exhibitions (capstone projects, Certificates of Initial Mastery, and senior projects) to serve as summary assessments of student mastery of a school's PBGRs
- Portfolios to collect student performance evidence aligned with school PBGRs for both formative and summative assessment of student progress and documentation of student academic, personal, and career goals
- Common tasks to give students varied opportunities to demonstrate proficiency and give teachers information for use in guiding curriculum, instruction, and assessment
- Proficiency-based departmental end-of-course exams that are mapped to individual courses, PBGRs, and Grade-Level/Grade-Span Expectations.

In 2008, the Rhode Island Department of Education (RIDE) certified 30 districts or schools as meeting the RIDE requirements for the new diploma system. Approval of the remaining 8 districts or schools is pending further review (Rhode Island Department of Education, 2008).

Strengths

One of the primary strengths of Rhode Island's PBGRs in the arts is that the system includes multiple measures at the local (district and school) level that are aligned with state PBGRs and state and national standards for arts education. Although many features of the local assessment proficiency-based assessment system are still under development, students will potentially have a wide variety of opportunities to demonstrate proficiency in the arts through exhibitions, graduation portfolios, as well as end-of-course assessments and various other measures. This flexibility in demonstrations of arts proficiency is designed to avoid narrowing the high school curriculum to focus too much instructional time and effort on achievement in subjects included in the statewide accountability tests (ELA, mathematics, and science). Such flexibility in local assessment also supports the requirement for graduation that all students demonstrate proficiency in at least one art form. In addition, the system allows for opportunities to demonstrate proficiency in the arts at multiple levels, including levels of proficiency that all students can realistically attain and levels of proficiency that would prepare students to pursue postsecondary studies and careers in the arts (Rhode Island Department of Education, 2004, p. 25).

Limitations

Rhode Island still has much work to do to fully realize its plans for proficiency-based graduation requirements in the arts. Much work (currently being supported by the Rhode Island Arts Learning Network) is still needed to develop definitions of arts proficiency, proficiency-based graduation requirements, and plans for local districts to assess arts proficiencies. Ultimately, effective implementation of local specialized assessments in the arts will require high-level and intensive professional development and support at the state and local levels. Work in developing and implementing an aligned system of local and state PBGRs in the arts will clearly need to continue for many years to come.

SOUTH CAROLINA

The South Carolina Arts Assessment Program (SCAAP) is the first and only web-based large-scale arts assessment in the United States. SCAAP was initiated in 2000 as a collaboration of the South Carolina Department of Education, the University of South Carolina, and South Carolina arts educators. SCAAP began its development work with fourth-grade music and visual arts assessments aligned with the South Carolina Visual and Performing Arts Standards. Since 2003, these assessments have been used as part of the evaluation of schools and districts receiving South Carolina Distinguished Arts Program grants. More recently, SCAAP has piloted dance and theater assessments intended for use with entry-level students in these subjects in middle school or high school. The long-term objective of the SCAAP is to develop separate assessments in all four arts disciplines that will be available for statewide use at several grade levels (Yap, Pearsall, & Peng, 2007). (Background information about South Carolina Arts Assessment Program in this section was obtained from the SCAAP website at http://scaap.ed.sc.edu.)

SCAAP assessments in all four arts disciplines combine web-based multiple-choice items and performance tasks. The SCAAP music assessment and the SCAAP visual arts assessment each include two 45-item multiple-choice test forms and two performance tasks. Because the SCAAP assessments are administered online they can include multimedia materials such as digital images, videos, and audio files. This is an unusual design feature for multiple-choice assessment items and opens up the possibility for inclusion of a wider range of authentic arts materials in the assessment. Audio or visual materials may be included as part of the multiple-choice item prompt and/or as part

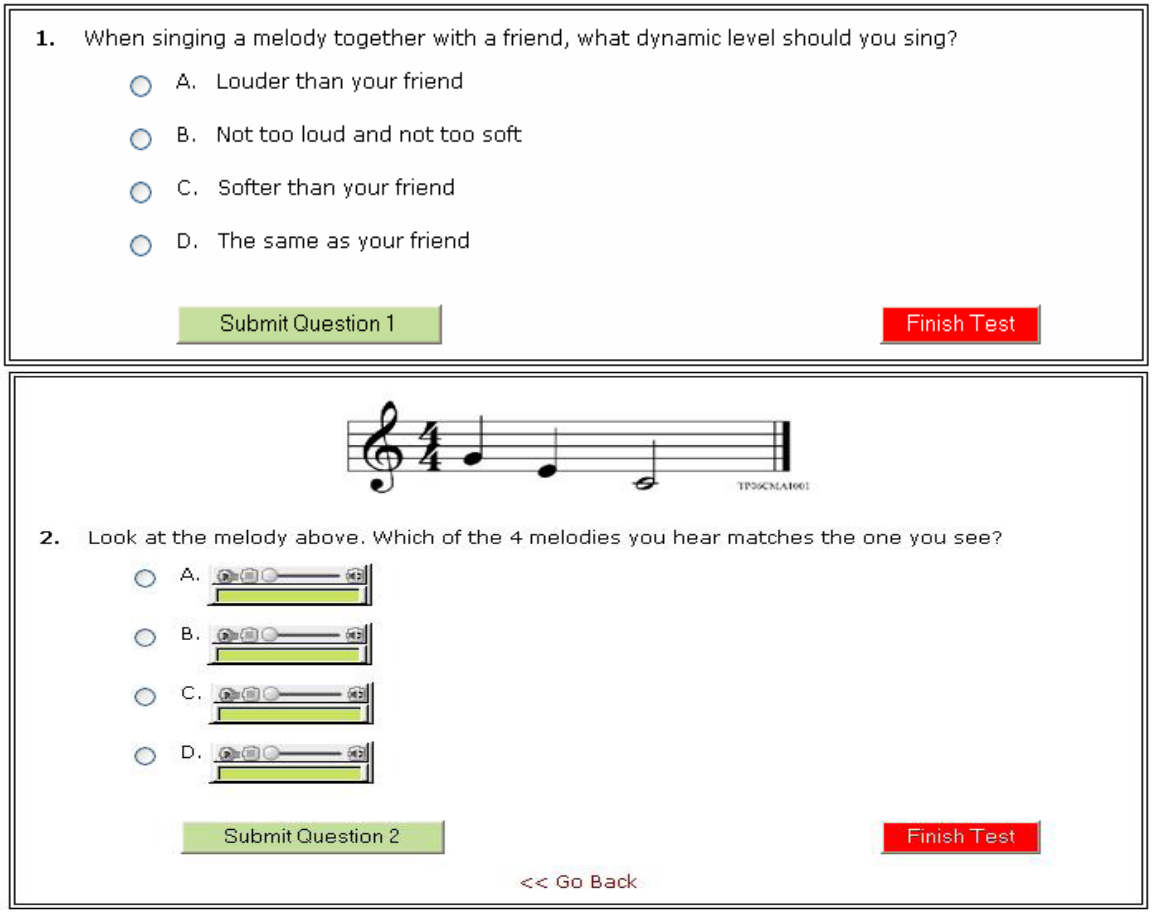
of one or more of the four response options. In addition to the 45-item multiple-choice test, students taking a SCAAP assessment are given two performance tasks. The performance tasks are administered individually and require students to demonstrate skills described in state standards. Examples of music performance tasks include performing a familiar tune on a neutral syllable ("du") and performing a rhythm improvisation using rhythm syllables (Yap, Lewis, & Feldon, 2007).

One of the two parallel multiple-choice test forms is administered to all fourth-grade students in participating schools. The two performance tasks in music or visual arts are administered to a sample of 50 students in each participating school and 150 students from each participating district. Multiple-choice questions are administered directly through the website (see Exhibit 4 for sample practice items from the fourth-grade music assessment), while responses to performance tasks (visual artworks and music performances) are digitally recorded online or scanned and uploaded to the website.

SCAAP multiple-choice test items are scored automatically as responses are entered online. Online submission of responses to performance tasks allows for remote scoring by a group of trained raters. Arts experts are recruited and trained as raters. Rater training includes use of selected exemplars of student performance to anchor ratings and web-based monitoring of scoring by raters in training to ensure high levels of interrater reliability. Raters must pass a qualifying test that compares their scoring of 15 randomly generated items with that of a validation committee (Yap, Lewis, & Feldon, 2007). Scoring of performance tasks makes use of item-specific analytic and holistic rubrics. Results are made available to the teacher online.

Use of the SCAAP music and visual arts assessments has so far been restricted to schools and districts that received state arts grants, but the numbers of participating schools and students have increased year by year, with 12 schools and approximately 1,700 students having participated in 2002 and 77 schools and just under 6,000 students in 2006.

Exhibit 4
Sample Practice Items from the SCAAP Fourth-Grade Music Assessment



Source: South Carolina Arts Assessment Program (SCAAP), Test Administration Manual 2007: Music. South Carolina Arts Assessment Program, Office of Program Evaluation, University of South Carolina, Columbia, SC

Strengths

The unique strengths of the SCAAP assessment are related directly to the system's uses of technology. The combination of web-based delivery of multimedia multiple-choice items and the online collection of responses to performance tasks are what set the SCAAP assessments apart from other standards-based arts assessments. As previously noted, adding multimedia capabilities to multiple-choice test items increased the authenticity of these item formats for arts assessment. The fact that assessments are administered online and then scored automatically in the case of multiple-choice items or by a centralized group of qualified raters in the case of performance tasks permits a high degree of standardization and reliability in test administration and scoring. Online administration, scoring, and web support for rater training are all features of the SCAAP that reduce the cost of scaling up the testing program for larger scale use. Other states have expressed interest in adapting the SCAAP approach, most notably Florida, which is currently pursuing development of an online music assessment following the SCAAP model (T. Pearsall, personal communication, January 17, 2008).

Limitations

Technology is also the chief limiting factor for the SCAAP. Schools and students must have access to networked computer workstations for the test administration. The development of online delivery systems for the assessment items, the digital recording of assessment responses, and the development of online reporting systems make the cost of creating and maintaining the SCAAP relatively high (though, to a degree, added development and maintenance costs may be offset by lower costs for revising and updating test forms relative to paper and pencil formats). Finally, the SCAAP is not truly a statewide arts assessment. Its current use is restricted to schools that are South Carolina Distinguished Arts Program grant recipients.

MINNESOTA

The arts are a required component of K–12 education in Minnesota, and students must fulfill an arts requirement to graduate. Legislation mandates, however, that the state cannot create statewide assessment in the arts (Minnesota Statute Section 120B.30), and although schools must report how they are meeting the standards there is no follow-up to ensure that schools are compliant in providing arts instruction (P. Paulson, personal communication, December 13, 2007). In the absence of a state-level accountability or assessment system in the arts, Minnesota relies on school-based arts assessment and accountability and uses a state-level arts education agency, the Perpich Center, to provide programs for improving arts instruction throughout the state. The Perpich Center has adapted a model for school-based arts assessment and accountability originally developed in Queensland, Australia.

In the Queensland system, graduation assessments are designed by individual teachers or schools but must align with state-level standards and criteria if the school wants to be accredited in the discipline. Schools accomplish this by developing a work plan in each subject area that is based on a state syllabus. In addition to approving the school's work plan, a panel at the state level also evaluates the teacher's assessment program. At the secondary school level, teachers use formative assessments, which can accumulate toward a final summative assessment. Exhibit 5 provides an example of a formative eighth-grade drama assessment from Queensland. When students reach the senior level, the teacher assesses their accumulated projects for graduation. Teachers' assessments are then evaluated through a moderation process, in which a panel at the state level verifies their judgments about student achievement. Schools are responsible for providing their assessment programs to the state board, along with sample student portfolios at each scoring level. At the state level, a panel determines whether or not the school has made appropriate judgments about student work by aligning the work with standards-based criteria. If the panel does not agree with the school, it consults with the school until agreement is reached. Each year, the state collects a random sample of moderated student portfolios to determine the degree of agreement between schools and the state panel. The random sampling process is used to confirm that the moderation process is effective at evaluating student achievement (Queensland Studies Authority, 2005).

Exhibit 5
Sample Assessment Item for Year 8 Drama from a Queensland School

In the following example, the teacher created an assessment document outlining the standards that the student is expected to demonstrate, and how, and the student must sign the document to show that he/she understands what is expected. The teacher then devised a form for assessing student performance in that standard.

DR 4.1 Students select dramatic elements and conventions to collaboratively shape improvisations and role-plays.
Opportunities to demonstrate this outcome will be given regularly as part of everyday classroom activities. Your teacher will observe and note your work during a number of improvisations, role-plays and drama games. You will be given opportunities to apply choices about dramatic elements such as role, relationships, movement, focus, mood and symbol through your class workshops, process dramas and a formal improvisation assessment task. Your teacher will keep a checklist that documents your progress.

I have read this assessment overview and understand what I am required to do this semester.

Name: _____

Signature: _____ Date: _____

Drama Outcome 4.1

Able to create a character profile from published script (see workbook).	Roles continue to develop the drama and stay true to their intent.	Relationships are believably created.	Tension of relationship has been carefully thought out and resolution is present and believable.	Action is focused in both place and time and connects to the rest of the drama.	Students gestures are carefully managed to symbolise character.	Selects specific objects, props or costumes as symbols to focus the action, enhance the mood and convey character.

Comments:

Source: http://www.learningplace.com.au/deliver/content.asp?pid=17949, curriculum and assessment developed by teachers at Bundamba State Secondary College

The Perpich Center adapted the Queensland assessment process for Minnesota by creating a panel to give feedback on teachers' locally designed assessment. They created a system in which teachers would serve on the assessment panel in rotations, and the assessment panel would be trained in standards-based arts assessment. The teachers initially selected to serve on the panel were those who had already been involved at the state level, writing arts standards and conducting professional development. They were enlisted to collect samples of student work and develop rubrics as examples that teachers could adapt for use in their classroom. In this process, they found that there were significant disagreements about what was critical to fair and reliable assessment of student artwork (P. Paulson, personal communication, December 13, 2007). For example, the teachers on

the panel had different ideas about what a body of work should consist of and different interpretations of key concepts and terms in the standards. Before developing and piloting rubrics, they had to come to agreement about what the standards were asking for and what student work was needed to demonstrate achievement.

To pilot the rubrics on a larger scale, the Perpich Center partnered with school districts from different parts of the state to check each other's assessment process. School districts would score their own students' work and then send it to the partner district to be scored. In this process the Perpich Center found that teachers did not understand the standards and were not developing sequential curricula. After the first pilot round, the Perpich Center started its work with subsequent partner districts by analyzing the standards with them and helping them with curriculum mapping. The process that was originally intended to provide teachers with assessment tools and guidance also became professional development on standards-based curriculum and instruction. Additionally, in partnering with districts from across the state, the Perpich Center created a peer network in which teachers share ideas about how to develop a standards-aligned curriculum and assess student achievement (both arts specialists and classroom teachers participate in the network). In its current state, the panel assessment process is not yet used for grading individual students but is used to help teachers develop consistent tools and knowledge about standards-based arts teaching and assessment (P. Paulson, personal communication, December 13, 2007).

Strengths

Despite differences in context, there are some valuable insights to be gained from the school-based assessment model developed in Queensland. Like the classroom-based model, it is an approach that facilitates performance assessment on a large scale. This model, however, gives teachers more control over the instruction and assessment program and develops their own capacity to provide standards-based arts instruction. Reports indicate that instruction and curricula in the arts have improved with the school-based assessment system. Because teachers are accountable for developing standards-based and criteria-aligned curricula and assessments, their capacity to do so has increased (Beattie, 1997).

The potential of adapting the school-based assessment system to the United States is demonstrated by the Quality Teacher Network in Minnesota. Although it is still in development as a system for assessing student achievement, the QTN has functioned as a professional development process for arts teachers in Minnesota and has moved the state toward more consistent understanding of and instruction with the arts standards. According to Pam Paulson from the Perpich Center (personal communication, December 13, 2007), it has helped teachers to see what is going on in their students' work and has helped them understand what forms of student work are needed, in addition to performance and portfolio pieces, to understand their arts learning. Furthermore, by establishing a peer panel to provide assessment tools and feedback and networking across districts, teachers are actively engaged in learning how to adapt their instructional practices to help their students meet the standards.

Limitations

The school-based assessment system in Queensland was born out of an educational environment very different from that currently in the United States. When the Queensland system was created, it was a movement away from an external examinations model toward a model that gave teachers autonomy in developing an instructional program. It is also very different from the U.S. system in that the purpose of state oversight is to accredit secondary schools to provide instruction in the subject. In Queensland, teachers did not take easily to the responsibility of developing and administering assessments (Childress, 2006), suggesting that implementation would require time and professional development for teachers to develop arts assessments that are standards aligned and reliable.

IMPLICATIONS FOR POLICY AND PRACTICE

In a recent report addressing the question of whether or not the public interest is being served by accountability and assessment in K–12 education, Joan Herman concludes that while the answer is of course complicated, available evidence suggests that the basic theory of action underlying accountability is generally working. In her view, that basic theory of action can be summed up as follows: "Accountability systems make public expectations and motivate educators and students to pay attention to learning and performance: Schools are changing what they are doing, they are focusing on teaching and learning and aligning curriculum and instruction with standards—*or at least those that are tested*" (Herman, 2007, p. 18, emphasis added). Does Herman's conclusion that accountability testing drives teaching and learning also apply to the arts? Several of the key findings of the research reviewed by Herman (2007, pp. 7–8) are worth noting in considering the desirability and potential of large-scale assessment and accountability in supporting implementation of standards-based arts education:

- State assessments focus instruction
- Teachers model what is assessed
- Schools focus on the test rather than the standards
- What is not tested becomes invisible.

Research on the effects of large-scale accountability testing on what is taught and learned in K–12 schooling (see essays collected in Herman & Haertel, 2005) highlights both the positive and the negative potential impacts of test-based accountability. On the positive side, research has shown that assessment for the purpose of accountability can have positive affects on the alignment of instruction with standards (that are tested). However, the negative impacts of current assessment-driven accountability policies are also clear. In a recent article enumerating the design flaws in NCLB accountability provisions, Richard Elmore (2003, p. 6) made the following observations: "Standardized testing is relatively cheap and easy to implement. Capacity building is expensive and complex. Policymakers generally like solutions that are simple and cheap rather than those that are complex and expensive. When we bear down on testing without the reciprocal supply of capacity, however, we exacerbate the problem we are trying to fix."

Approaches to accountability for arts education that focus on monitoring resource inputs and opportunities to learn and that do not rely primarily on standards and assessment of student learning are certainly possible (see Darling-Hammond, 2004) and may be desirable given the current state of testing fatigue in the wake of NCLB and the technical and practical challenges posed by developing and implementing large-scale arts assessment systems. However, the power of standards-based testing to shape the delivered curriculum should not be too quickly dismissed. Standards coupled with standards-based assessments can play important roles in supporting the alignment of individual and collective expectations within schools and thus foster the internal accountability that leads to increased opportunities to learn and improved learning outcomes. In this light, the key to designing an effective system of arts education assessment and accountability—a system that supports expanded student access to sequential, standards-based arts instruction and leads to higher levels of student proficiency in the arts—is the alignment of internal (within school) and external (district, state, and national level) accountability with common definitions of learning goals for the arts (standards) and effective assessments of those goals for multiple purposes (to inform learning and instruction as well as to communicate learning results externally). This alignment requires consensus on high-quality standards as well as high-quality assessments.

Creating quality and consensus in education standards and maintaining that quality over time are largely matters of constant, broadly-participatory, iterative development, review, and redevelopment. As anyone who has participated in standards development can attest, the work is grueling, contentious, and never fully done. Though recent reviews of the quality of state standards in other subject areas have not given standards high marks on key criteria such as extent and reasonableness of scope of content addressed, specificity, rigor, and consistency with sound models of learning (see, for example, Wilson & Berenthal, 2005), the National Standards for the Arts and the state arts standards are predominantly accepted as foundations for assessment and instruction in the arts.

Quality in standards is a critical starting point because quality in any assessment is always first and foremost a matter of being clear about what is to be assessed (referred to by test designers as the "construct"; see Messick, 1994). In theory, standards provide the targets for both instruction and assessment. In practice, as noted in Herman's (2007) review, assessments may usurp the role of standards in communicating key learning goals. This reality is particularly problematic in the arts when traditional assessment approaches are used. Traditional testing approaches can result in the content of assessments being defined too narrowly. In these cases, test results may capture only a small part of the important learning that has taken place in the arts. This narrow approach to assessment can have many unfortunate consequences: the dissatisfaction of teachers and artists, the tendency to teach to a narrow set of arts goals, and the rejection of standards-based assessment altogether. For example, in commenting on the limitations of the AP studio arts criteria, Lois Hetland (personal communication, June 15, 2008) noted, "It's easiest to assess craft, and that's what AP criteria do. But that's only a proxy for what we want in a developed artistic thinker/maker; if getting high results on the outcome measures does shape the programs away from the development of creative thinking, connection making, metaphorical thinking, explorations, and so forth, then it's destructive to arts education."

Simply adopting a performance assessment approach alone will not guarantee that the target of assessment is a valued learning goal. Nonetheless, there are clear advantages of using performance assessment methods for assessing important learning goals in the arts. As noted in *The Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999, p. 137, emphasis added), "[e]very test, regardless of its format, measures test-taker performance in a specified domain. *Performance assessments, however, attempt to emulate the context or conditions in which the intended knowledge or skills are actually applied*." Emulating the conditions in which artistic proficiency is applied is the key strength of performance assessment for use as a measure of student learning in the arts.

Beyond issues of the quality and appropriateness of assessment formats (traditional versus performance based) it is important to consider the quality of assessment systems. Effective accountability assessment for arts education will require the application of different types of assessments for multiple purposes. In *Knowing What Students Know: The Science and Design of Educational Assessment*, the National Research Council (2001) identified the features that characterize an ideally balanced educational assessment system: comprehensiveness, coherence, and continuity. For an assessment system to be comprehensive means that "a range of measurement approaches should be used to provide a variety of evidence to support educational decision-making" (National Research Council, 2001, p. 238). To be coherent, there must be a strong link between assessments at the school, district, and state levels (vertical alignment) and between curriculum, instruction, and assessment (horizontal alignment)—all should be aligned with common definitions of proficiency and models of learning (p. 240). Continuity in an assessment system relates to the ability of the system to track learning progress over time.

None of the existing large-scale arts assessment models or statewide arts assessment systems reviewed in this report meet all the criteria of a completely balanced educational assessment system. Yet, when considered against the criteria for a balanced assessment system, each assessment approach highlights elements of effective assessment system design that may be useful in the design of an arts assessment system in California. Exhibit 6 summarizes the degree to which each of the large-scale arts assessment and state arts assessment systems discussed in this report meet the criteria for a balanced assessment system.

Exhibit 6
Balance in Large-Scale Assessment Systems

	Comprehensiveness	Vertical Alignment (Coherence)	Horizontal Alignment (Coherence)	Continuity
1997 NAEP Arts Assessment	Within the constraints of a one-time, on-demand assessment, the combination of creating/performing and responding items and the use of matrix sampling permit a relatively high level of variation in assessment formats and content.	Alignment of the NAEP Arts Assessment framework with National Arts Standards is a good model of vertical alignment, but effective vertical alignment of school, district, and state standards within a state-level assessment would require greater specificity of standards and broader coverage of standards by assessments.	The NAEP design and purpose are not compatible with alignment of assessments with curriculum and instruction.	By design, NAEP lacks the ability to monitor individual learning trajectories because the purpose is to provide a population profile of proficiency at a single point of time.
IB arts portfolio exam	As a portfolio assessment, the IB allows for inclusion of a variety of assessment formats.	Vertical alignment is strong within the IB program but is made possible by a level of organizational support (local, national, and international) that public school systems will find hard to duplicate.	Curricular guidelines and monitoring by a central Curriculum Office give IB schools a degree of flexibility while at the same time ensuring strong alignment of curriculum, instruction, and assessments.	Inclusion in the portfolio of multiple products, including best finished works as well as works in progress provides a good window on learning and achievement over time (though within a limited time frame) .
AP Studio Art exam	As a set of portfolio assessments, the AP Studio Art exam allows for inclusion of a variety of assessment formats.	Vertical alignment is not a goal of the AP Studio Art exam.	Horizontal alignment is achieved by "teaching to the test" and design of curriculum and instruction to align with AP portfolio guidelines.	Like the IB portfolio, the AP portfolios capture important aspects of artistic development, though within a very limited time frame.
Kentucky	Paper-and-pencil format of the KCCT and limited number of items restricts range of measures possible. Only 3 of 5 subdomains of state arts standards are covered.	Vertical alignment is limited by the small number of arts items on the KCCT and the traditional item format, both of which restrict content coverage.	Limited content and grade level coverage of the KCCT also limit alignment with curriculum and instruction.	Limited content coverage, testing in grades 5, 8, and 11 only, and traditional item formats restrict ability to monitor individual learning over time.
Washington	Performance assessment format of CBPA and embedding of assessment within instruction allow for great variety in types of measures.	Fact that CBPA are developed at the state level and are delivered as part of classroom instruction supports strong vertical alignment.	Fact that CBPA are arts activities integrated with curriculum supports strong horizontal alignment.	Capacity of the CBPA to monitor individual learning over time appears to be strong but has not yet been demonstrated.

	Comprehensiveness	Vertical Alignment (Coherence)	Horizontal Alignment (Coherence)	Continuity
Rhode Island	Includes a great variety of assessment types and multiple opportunities to demonstrate proficiency.	Vertical alignment will depend on the alignment of PBGRs at local and state level and ability to align local and state assessments to common PBGRs.	Integration of curriculum, instruction, and assessment is strong suit of the local proficiency-based assessment systems being developed in RI.	PBGRs at state and local level are basis for aligning multiple measures of proficiency over time (through high school years).
South Carolina	Combination of traditional item formats enhanced with multimedia plus performance tasks allows for great variety of measures.	Alignment of assessments with South Carolina arts standards is a strength, but influence of standards and SCAAP assessments on district and school level assessment is not known.	Potential for alignment of school curriculum and instruction with SCAAP assessment is strengthened by teacher professional development but limited by use of SCAAP only in schools receiving SC Distinguished Arts Program grants.	SCAAP assessment is currently only used in grade 4 (music and visual arts) and with entry level arts students in middle and high school (theater and dance). Monitoring individual learning progress is not possible.
Minnesota	School-based assessment development creates the potential for a wide variety of measures, but variety may not be present in every school. Professional development of arts teachers adds to potential comprehensiveness of assessment.	As in California, Minnesota state law prohibits use of a statewide assessment in the arts. However, state panels to evaluate teacher assessment plans and the Arts Quality Teacher Network contribute to a degree of potential vertical alignment.	Locally designed assessment enhances alignment with school curriculum and instruction. Moderation process of local assessments by state panels also contributes to horizontal alignment.	School-based assessment development strengthens potential for effective monitoring of individual learning over time.

As discussed above and summarized in Exhibit 6, none of the large-scale arts assessments or state arts assessments we have reviewed meets all the criteria for a perfectly balanced assessment system. However, the large-scale arts assessment models, state arts assessments, and related literature discussed in this review offer many lessons for the design of state arts assessment systems. Among these lessons are the following:

- Whenever assessment results are used to support high-stakes decisions (as in school or district accountability), multiple measures and a variety of assessment measures are inherently superior to any single assessment format. The purposes of accountability and instructional improvement are best served by a broad range of evidence of student learning.
- Performance assessment aligned with standards is desirable but costly and technically challenging for large-scale development and implementation. When possible, matrix sampling (as in the NAEP) can reduce time and costs associated with administration of extended performance tasks on a large scale. Adopting a hybrid assessment approach (as in South Carolina's SCAAP) that combines selected response (multiple choice) and performance assessment items offers some of the advantages of both testing formats.
- Portfolio assessments with clear guidelines for selecting and evaluating student work products (such as the IB arts portfolio assessment and the AP Studio Art portfolios) are desirable because they present clear goals for instruction and multiple opportunities to assess student growth and achievement in the arts. When portfolio assessments are embedded in the learning process and are combined with other performance assessments (as they are in Washington's CBPA and Rhode Island's proficiency-based assessments), they are excellent tools for monitoring student learning and guiding arts instruction over extended periods of time.
- Use of online and digital media (as in South Carolina's SCAAP) can broaden the range of stimulus and response options in arts assessment and can also reduce time required and raise levels of standardization in test administration and scoring.
- Because of the opportunities it provides for teacher learning and professional development, a school-based assessment system such as that developed in Australia and being adapted for use in Minnesota may have strong potential for developing school-level capacity to implement high quality standards-based sequential arts instruction.
- A coordinated system of local proficiency-based assessments (relying on multiple measures), alignment of state and local definitions of proficiency in the arts, and a requirement to demonstrate proficiency in the arts for high school graduation (as in Rhode Island) provides a model for vertical and horizontal alignment of arts assessments and may build capacity for expanded implementation of standards-based arts education.
- Given the technical challenges posed by assessment of student learning in the arts, state-level professional development and moderation of the quality of local (school or district level) assessments will be essential elements in support of effective large-scale arts assessment and accountability

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *The standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Baker, E. L. (1998). *Model-based performance assessment*. CSE Technical Report 465. Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Barton, P. E. (1999). *Too much testing of the wrong kind; too little of the right kind in K–12 education*. Princeton, NJ: Educational Testing Service.

Beattie, D. K. (1997). Visual arts criteria, objectives and standards: A revisit. *Studies in Art Education*, 38(4): 217–231.

Bodilly, S. J., Augustine, C. H., & Zakaras, L. (2008). *Revitalizing arts education through community-wide coordination*. Santa Monica, CA: RAND Corporation.

Bonbright, J. N., & McGreevy-Nichols, S. (1999). NAEP and dance: On contextual data, process, and problems in dance assessments, and recommendations for the field. *Arts Education Policy Review, 100*(6): 27–32.

Boughton, D. (2004). Assessing art learning in changing contexts: High-stakes accountability, international standards and changing conceptions of artistic development. In E. W. Eisner & M. D. Day (Eds.), *Handbook of Research and Policy in Art Education*. Mahway, NJ: Lawrence Erlbaum.

Childress, J. (2006). Conversations…with Dr. Donna Kay Beattie. *New York State Art Teacher Association (NYSATA) News*. Retrieved on August 7, 2008, from http://www.nysata.org/doc/pdf/DK_Beattie_Interview.pdf

College Board, (2006). *AP report to the nation*. Princeton, NJ: College Board.

College Board. (2007a). *AP Studio Art course description: May 2008*. Princeton, NJ: College Board. Retrieved on August 7, 2008 from http://www.collegeboard.com/student/testing/ap/sub_studioart.html

College Board. (2007b). *AP Studio Art 2007 scoring guidelines*. Princeton, NJ: College Board. Retrieved on August 7, 2008 from http://www.collegeboard.com/student/testing/ap/sub_studioart.html

Colwell, R. (1999). The 1997 assessment in music: Red flags in the sunset. *Arts Education Policy Review, 100*(6): 33–39.

Colwell, R. (2003). The status of arts assessment: Examples from music. *Arts Education Policy Review, 105*(2): 11–18.

Consortium of National Arts Education Associations. (1994). *National standards for arts education: What every young American should know and be able to do in the arts*. Reston, VA: Music Educators National Conference.

Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review 64*(1), 5–30.

Darling-Hammond, L. (2004). Standards, accountability, and school reform. *Teachers College Record, 106*(6), 1047–1085.

Darling-Hammond, L., & Rustique-Forrester, E. (2005). The consequences of student testing for teaching and teacher quality. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountabilit*y, pp. 289–319. Malden, MA: Blackwell.

Donmoyer, R. (1999). Rendering unto Caesar that which is Caesar's: A critique of the NAEP in theatre. *Arts Education Policy Review 100*(6): 21–26.

Eisner, E. W. (1999). The National Assessment in the visual arts. *Arts Education Policy Review 100*(6):16–20.

Elmore, R. (2003). A plea for strong practice. *Educational Leadership 61*(3): 6–10.

GAO. (2003). Title I: Characteristics of tests will influence expenses; information sharing may help states realize efficiencies (GAO-03-389). Washington, DC: United States General Accounting Office.

Haertel, E. H., & Herman, J. L. (2005). A historical perspective on validity requirements for accountability testing. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountabilit*y, pp. 1–34. Malden, MA: Blackwell.

Herman, J. L. (2004). The effects of testing on instruction. In S. H. Fuhrman & R. F. Elmore, (Eds.), Redesigning accountability systems for education, pp. 141–166. New York, NY: Teachers College Press.

Herman, J. L. (2007). *Accountability and assessment: Is public interest in K–12 education being served? CRESST Report 728.* Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education and Information Studies, University of California, Los Angeles.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment.* Alexandria, VA: Association of Supervision and Curriculum Development.

Herman, J. L., & Haertel, E. H. (Eds.). (2005). *Uses and misuses of data for educational accountability. The 104 yearbook for the National Society for the Study of Education*, Part 2. Malden, MA: Blackwell.

Horn, D. (2005). *Status of arts education in Kentucky public schools.* KY: The Collaborative for Teaching and Learning.

Illinois Creates. (nd). *Arts at the core: Every school, every student.* Retrieved on February 1, 2008, from http://www.artsalliance.org/docs/education/artsAtTheCore.pdf

Joseph, A. (2006). *Arts assessments for Washington State Classroom-Based Performance Assessments (CBPAs): The journey in progress.* Retrieved on August 7, 2008, from http://www.k12.wa.us/curriculumInstruct/arts

Kentucky Department of Education. (n.d.). *Kentucky revised statutes*. Retrieved on June 26, 2008, from http://www.kde.state.ky.us/KDE/Administrative+Resources/Testing+and+Reporting+/CATS/Policies+and+Regulations/

Kentucky Department of Education. (2006).*Core content for arts and humanities assessment, version 4.1*. Retrieved on June 26, 2008, from http://www.kde.state.ky.us/KDE/Instructional+Resources/Curriculum+Documents+and+Resources/Core+Content+for+Assessment/

Kentucky Department of Education. (2007). *Blueprint for Kentucky Core Content Test.* Retrieved on June 26, 2008, from http://www.education.ky.gov/KDE/Administrative+Resources/Testing+and+Reporting+/CATS/Accountability+System/Blueprint+for+Kentucky+Core+Content+Test.htm

Linn, R. L. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis, 15*(1), 1–16.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15–21.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher 23*(2), 13–23.

Mishook, J. L., & Kornhaber, M. L. (2006). Arts integration in an era of accountability. *Arts Education Policy Review 107*(4): 3–11.

Morton, C. (1999). A national arts test: An overview of the NAEP Report. *Teaching Theatre 10*(3): 3–6.

Myford, C. M., & Sims-Gunzenhauser, A. (2004). The evolution of large-scale assessment programs in the visual arts. In E. W. Eisner & M. D. Day (Eds.), *Handbook of research and policy in art education*. Mahway, NJ: Lawrence Erlbaum.

National Assessment Governing Board. (1994). *1997 arts education assessment framework.* Washington, DC: National Assessment Governing Board.

National Center for Education Statistics. (1996). *Technical issues in large-scale performance assessment*. Washington, DC: U. S. Department of Education.

National Commission on Educational Excellence. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U. S. Department of Education.

National Commission on Testing and Public Policy. (1990). *From gatekeeper to gateway: Transforming testing in America*. Chestnut Hill, MA: Author.

National Council on Educational Standards and Testing. (1992). *Raising standards for American education: A report to Congress, the Secretary of Education, the National Education Goals Panel, and the American people*. Washington, DC: U. S. Government Printing Office.

National Education Goals Panel. (1994). *The 1994 national education goals report: Building a nation of learners*. Washington, DC: Author.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment, J. W. Pellegrino, N. Chudowsky, and R. Glaser (Eds.), Center for Education. Division on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

New York State Education Department. (2001). *High school (9th grade) arts assessment sampler draft*. Albany, NY: The University of the State of New York, State Education Department. Retrieved on August 7, 2008, from http://www.emsc.nysed.gov/osa/arths/hssampart.html

Pedersen, P. V. (2007). What is measured is treasured: The impact of the No Child Left Behind Act on nonassessed subjects. *The Clearing House 80* (6): 287–291.

Persky, H. (2004). The NAEP Arts Assessment: Pushing boundaries of large-scale performance assessment. In E. W. Eisner & M. D. Day (Eds.), *Handbook of research and policy in art education*. Mahway, NJ: Lawrence Erlbaum.

Persky, H., Sandene, B. A., & Askew, J. (1998). The NAEP 1997 arts report card (NCES 1999–486). Washington, DC: National Center for Education Statistics.

Pitman, J. A. and Dudley, R. P. (1985, republished in 1998). *Criteria-based assessment: The Queensland experience.* Spring Hill, Australia: Queensland Board of Senior Secondary School Studies. Retrieved on August 7, 2008 from http://www.qsa.qld.edu.au/downloads/publications/research_qbssss_assessment_criteria_98.pdf

Pistone, N. (2002) *Envisioning arts assessment: A process guide for assessing arts education in school districts and states.* Washington, DC: Arts Education Partnership and Council of Chief State School Officers.

Queensland Studies Authority. (2005). Moderation Processes for Senior Certification. Spring Hill, Australia: Queensland Government. Retrieved on August 7, 2008 from, http://www.qsa.qld.edu.au/downloads/assessment/assess_snr_moderation.pdf

Resnick, L. B., & Resnick, D. (1992). Assessing the thinking curriculum: New tools for education reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 301–328). Boston: Kluwer Academic Publishers.

Rhode Island Department of Education. (2004). *Initial guidance for the graduation by proficiency component of the regulations of the Board of Regents for Elementary and Secondary Education regarding public high schools and ensuring literacy for students entering high school*. Providence, RI: Rhode Island Department of Education.

Rhode Island Department of Education. (2005). *Rhode Island's diploma system: An overview*. Providence, RI: Rhode Island Department of Education.

Rhode Island Department of Education. (2008). Preliminary approval for new diploma systems awarded to 30 schools, districts. RIDE News Release, January 31, 2008.

Ross, J. (1994). The National Standards for Arts: The emperor's new clothes. *Arts Education Policy Review 96*(2):26–30.

Schneider, M. C. (2003). Merging the ideal with reality: Reflecting on the NAEP 1997 Arts Assessment and looking toward 2008. *Arts Education Policy Review 104*(4): 17–22.

Schultz, R. A. (2002). Apples, oranges, and assessment. *Arts Education Policy Review 103*(3): 11–16.

Soep, E. (2004). Assessment and visual arts. In E. W. Eisner & M. D. Day (Eds.), *Handbook of research and policy in art education*. Mahway, NJ: Lawrence Erlbaum.

State Collaborative on Assessment and Student Standards Arts Education Consortium. (1999). *Arts assessment: Lessons learned from developing performance tasks*. Washington, DC: Council of Chief State School Officers.

Washington State Arts Commission. (2006). *Arts for every student: Arts education resources initiative.* Olympia, WA: Washington State Arts Commission.

Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. New York: Jossey-Bass.

Wilson, M. R., & Berenthal, M. W. (Eds.). (2005). *Systems for state science assessment*. Washington, DC: National Academies Press.

Yan, J., & Rieder, S. (2001). *Maryland assessment of fine arts education: State-of-the-art in large-scale fine arts assessments.* Rockville, MD: Westat.

Yap, C. C., Pearsall, T., & Peng, P. (2007, April). A web-based arts assessment: Development & implementation. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Yap, C. C., Lewis, A., & Feldon, D. (2007, April). Efficacy of a web-based training and monitoring procedure in scoring performance tasks. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, IL.

SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Phone: 650.859.2000

www.sri.com