

MUSIC
EDUCATORS
NATIONAL
CONFERENCE



New England Conservatory
Music Education
290 Huntington Avenue
Boston, MA 02115

Number 21

Summer 1998

Special Research Interest Group In Measurement and Evaluation

TABLE OF CONTENTS

Introduction	1
State Arts Assessment	5
Portfolios: Assessment, Pedagogy, or Staff Development?	13
Authentically Assessing the National Content Standards in Music	18
Music Assessment Today	21

INTRODUCTION
RICHARD COLWELL
EDITOR
New England Conservatory

The SRIG session in Phoenix was again devoted to issues about the portfolio as an assessment device. The SRIG has an obligation to the profession to raise both technical and philosophical issues related to any and all assessment devices. The reform literature continues to imply that the portfolio or individual folder based measurement will improve instruction and improve our ability to report the results of instruction to parents, administrators, and the community. The research evidence for this, however, is missing. Arts Propel, a project conducted a decade ago in Pittsburgh, included music along with literature and visual arts, and is often cited as a model of portfolio evaluation. Preliminary inspection, however, indicates that the primary outcome in Pittsburgh was an instructional device, not an assessment tool. To provide the music education community with current information, MENC invited the primary music evaluator of Arts Propel to address the SRIG and to have two responses to his remarks, one from a college teacher educator, Dr. Patti Sink, and one from a K-6 music teacher, Dr. Brian Leavell.

To provide a beginning to the session, the chair passed out an article by Peter Berger, a Vermont middle school English teacher, who wrote an article entitled "Portfolio Folly" in the January 14th issue of *Education Week*. Berger states that if Vermont is leading the nation with portfolios, the news isn't good for the rest of the country. He suggests that portfolios are not new; Plato probably carried one to his Socrates lectures. He suggests that the rubrics are not meaningful; in Vermont portfolio scoring was:

REGIONAL EVALUATION REPRESENTATIVES

JAMES SHERBON - VICE CHAIR
UNIVERSITY OF NORTH CAROLINA AT
GREENSBORO

RICHARD SANG - EAST
QUEENS COLLEGE/ CUNY
AARON COPLAND SCHOOL OF MUSIC

VANCE WOLVERTON - WEST
CALIFORNIA STATE UNIVERSITY AT
FULLERTON

PATRICIA SINK - SOUTH
UNIVERSITY OF NORTH CAROLINA AT
GREENSBORO

BRIAN LEAVELL - SOUTHWEST
DENTON, TEXAS

EDITORS
RICHARD COLWELL & GREG GATIEN

frequently, sometimes, or rarely, a quantitative not a qualitative judgment.

As an introduction to this issue of the newsletter, some of the concerns about portfolios are articulated. Raising questions does not mean members of the SRIG are opposed to the use of portfolios. Quite the opposite. Portfolios bring a new dimension to assessment, a dimension that needs to be as carefully critiqued as have been standardized tests. We do believe that wholesale and exclusive adoption of the portfolio or any other device by officials of a state or large entity is misguided. Such an action resembles state mandates of phonics, Kodaly, or Microsoft. These individuals fail to understand the complexity of teaching and learning.

A few stated concerns or issues that have appeared in the literature and should be considered within the music education assessment community:

1. Vermont has had a major problem with obtaining adequate indicators of reliability despite intensive teacher training. Training that restricts the purview of the evaluation and the more creative uses of the portfolio could narrow instruction. Much of the training designed to have all judges look for a limited number of the same things is being done primarily to increase reliability indicators, not to improve the assessment. Probably, more damning in Vermont, is the lack of content validity. This is suggested by the fact that the evaluations of experts disagree with the evaluations of teachers more often than they agree. The definitions of validity and reliability of portfolios have changed, as we suggest later. Vermont has joined Kentucky in reducing the emphasis on portfolio assessment, especially in making any high stakes assessments.

2. Music teachers have found that students enjoy keeping recordings of their performances and comparing performances over time. Students can hear obvious changes but whether they can identify causal factors remains unknown. Making comparisons is an excellent self-

teaching device but does little to inform the student whether he or she is performing up to her potential or to any established performance standard.

3. Students involved with performance experiences have sent a clear message to all teachers that time spent in talking about music during ensemble practice is not time spent in making music. Students come to music class to sing and/or play; they are in ensembles for the cooperative experience of enjoyment more than any individual learning. Tapes and reflecting on them might best be done outside of class.

4. Feedback has been found to be a critical element in improving learning but the feedback needs to be immediate. Delayed feedback, such as that from a returned portfolio, has limited value for improving learning. Students deserve a quick, independent opinion on any work submitted for judgment.

5. Individuals developing portfolios have different definitions of reliability and validity from those traditionally taught in test and measurement courses. We all learned that a test can be reliable but not valid; a test can measure a concept consistently but the concept not be important. We also learned that a test could not be valid if it were not reliable. How could a test measure what a student has learned if one day the demonstration was rated outstanding and the next day the same demonstration rated a failure? The language of assessment now includes consequential validity which distinguishes between low and high stake decisions; the idea being that reliability is less important with low stakes assessment but still critical with high stakes evaluation. Developmental validity focuses on the appropriateness of the process.

6. National and state education policy reflect a continuing belief in the power of assessment to support and measure educational goals and this belief is more nearly a belief in high stakes outcomes than a formative or low stakes assessment. The public has never interfered

extensively in how instruction is conducted; the public's primary concern is with the results or outcomes of the instruction. Teachers are acknowledged experts on pedagogy; policy makers are the authorities on what outcomes can reasonably be expected of publicly supported instruction.

7. Of the nine voluntary national standards in music, skill is measured with fewer than half of the standards. There is authentic knowledge and authentic attitudes and feelings. Authentic exhibitions of knowledge are often oral or written. Just as one need not listen to an entire tape to judge a student's performance, one need not read an entire essay to judge whether a student has gained the essential knowledge and concepts from instruction. Sampling is authentic, short answers and recognition of errors such as one might find in a multiple choice test can also be authentic. An authentic listening experience based on instruction requires a cognitive map and may be a bit like audition. Listening is not always directly related to a student's singing or playing ability. Evaluating the adequacy of the readily deployable cognitive map can be accomplished efficaciously without the student spending critical "opportunity to learn time" on describing in writing, the map she is using and explaining why. The why is informative but an adequate sample is what allows for inferences.

8. Before one can suggest appropriate solutions to assessment dilemmas the problem must be identified. We can be relatively sure that the problem in music education is not excessive past reliance on multiple choice and true-false tests. Few assessments or grades in music have ever been based primarily on these one-shot tests. Grades are more often based on attendance or attitude than on any established paradigm from the testing and measurement field.

9. Linn, Baker, and Dunbar in 1991 suggested that the quality of an assessment be based on eight items:

consequences
 fairness
 transfer and generalizability
 cognitive complexity
 content quality
 content coverage
 meaningfulness
 cost and efficiency

Each of these should be considered in setting up a portfolio system.

10. Audio tapes are, of course, an incomplete assessment of performance and are especially weak with young students. Bad posture and poor embouchure formation are not readily apparent from an audio tape; some poor performance habits are not evident in a performance exhibition for years. There is also value in a student knowing cognitively the characteristics of a good embouchure, its value nearly equal to the ability to demonstrate a good embouchure. With knowledge, one can check oneself in the mirror, assist in cooperative learning, and begin to identify the physical causes of good and poor tone quality along with other knowledge based issues when listening to music.

11. Portfolios are not easy to score and would be especially troublesome if students and amateurs were to use them to derive deep meaning from the items contained therein. Most research has been conducted with writing portfolios and their success is possible because of intensive training of raters. This intensive training includes model or demonstration portfolios "appropriately" scored so that much of the individualism and external objectivity of the rater is lost. Although we cannot predict whether this "individualism" is a good or a bad thing, we do know that evaluators in music bring different perspectives to the task of judging adequacy; not only do they listen for different musical aspects but often bring a differing set of priorities and differing ideas about musical priorities based upon the age or developmental stage of the student. The writing portfolio has also been aided by specific instructions on what to include in the portfolio. When the "col-

lection" of items in the portfolio differs from student to student and class to class, the use of the portfolio as an assessment tool becomes more complex and often less reliable and less valid.

12. The purpose of the portfolio is different in various music education programs. We have attended in-service sessions where the portfolio was touted as the "magic bullet" for improving a methods course. These teacher educators were not attempting to correct any prior invalid assessment system in their course or bring about a more valid grading system for their class. The purpose was to bring more projects into the classroom, projects that could be reviewed and exhibited by one and all allowing a different type of reflection and criticism. This recommendation seems comparable to the long ex-tant student teaching log and systematic videotaping that occurs in student teaching and conducting classes.

13. A major issue in using the portfolio for assessment has been the adoption of rubrics. Again, the writing portfolio has taken a leadership role in demonstrating the importance of the language of rubrics. One project assessed writing samples on theme, character, setting, plot, communication, and narrative effectiveness. Music educators might be expected to establish comparable judgmental factors for portfolios. Even with accepted rubrics mean correlations across raters is likely to be low. In language arts the rubric "narrative effectiveness" was .45 as reported on page 12 "Issues in Portfolio Assessment," CSE Technical Report 410, 1996 by John Novak, Joan Herman and Maryl Gearhart.

14. Carol Myford and Robert Mislevy report in "Monitoring and Improving a Portfolio Assessment System" (MS 94-05) their work at ETS. Their concern was improving the evaluation of the advanced placement tests in visual arts. They stressed that one concern was the question of what is of value rather than what is correct. In this publication they point out two variables: The importance of student's knowing the assessment crite-

ria and the assessment complexity caused by the variety of art objects that students were allowed to place in their portfolios.

15. When students were asked to also comment on their works, these comments needed to be judged. They were assessed on clarity, coherence, and consistency. Thus, the ability to communicate is brought to the portfolio assessment task. Required writing has been found to raise the issue of fairness as the evaluator is no longer judging just the work of art (or musical performance) but judging the student's background, ability, and experience in expressing him or herself.

16. Myford and Mislevy are optimistic about judging works of art as they achieved decent reliability using a four point rating scale. They, however, provide scorers with three cautions: empathizing with the student based on cultural clues in the portfolio; the need to remain objective in rating a portfolio that follows an outstanding one (evaluator bounce); and the difficulty of not knowing the context in which the portfolio was prepared. (pp 20-21)

17. Recent ETS research has focused on developing a science portfolio which is exciting and demanding work. The researchers hope to judge student work and thinking in science that is consistent with the scientific ways of knowing, arguing, and exploring. ("Moving Toward a Portfolio Culture in Science Education" by Drew Gitomer and Richard Duschl, Center for Performance Assessment, MS #94-07) An explanation of suggested evaluation criteria asks students to respond to relationships - why and how things are alike; to clarify, to be consistent with evidence, to give examples, to make sense by predicting the outcome, and reporting on any surprises. In addition, the student is to acknowledge alternative explanations, elaborate on a theme and determine the accuracy of the student's own model with other models. (p 11) The emphasis in this portfolio exercise is truly evaluating critical thinking as the students must use syntactical knowledge to deal with rules and semantic knowledge to derive the content and meaning of the scientific principles.

18. In another Center publication, this one by Paul LeMahieu, Drew Gitomer and JoAnne Eresh (MS 94-01), the authors recommend the use of portfolios only when the purpose of the assessment is clear and teaching practices are consistent with the goal of the portfolio. (It is possible to determine how intense the teacher was in presenting the material being evaluated, raising the question of just who is being evaluated in the portfolio). In music, portfolios stressing performance is evidence of the student's home life and parental support.
19. LeMahieu et. al. suggest that there must be a shared interpretive framework by those using the assessment, there must be coherence in the school system (this will be an important question for music educators to address), and individuals within the system must discuss issues of quality in both instructional and psychometric terms. (p26)
20. Users should distinguish between score reliability and rater reliability, an issue that music educators have yet to think about.
21. There is a danger that music portfolios will focus on tasks and activities and not on the underlying basics of student learning. (p. 18, LeMahieu et. al.)
22. There are unexplained differences between ratings of middle school and high school portfolios. Also female judges give higher scores.
23. A reasonable guess is that some items within a portfolio will differ in their ease or difficulty in discrimination - some items will present a more difficult challenge in judging levels of proficiency. In the paper by LeMahieu, the suggested scoring rubrics consists of 7 items that define accomplishment in writing, four in use of processes and strategies for writing, and eight items defining growth, development, and engagement as a writer. The last rubric was the more difficult.
24. Research in England and Wales indicates that teachers must spend 2 to 3 hours of extra work daily when portfolios are incorporated into the curriculum. This amount of time is a high price to pay when the results are not minimally reliable. In the US, those states with the most experience with portfolios have decided that several assessment methods are better than one - there is a role for short answer, multiple choice, and other devices in assessing student outcomes.
25. A main concern is whether the test performance observed in a portfolio generalizes to a larger domain of knowledge and/or skills. The general perception in music education is that ability to perform well (the taped performance) does not generalize to all of the goals that music educators presently hold dear. Assessing all nine standards at several ages will require a variety of assessment methods, some untried.
26. One of the positive examples from the performance literature is that research by the military with large scale performance examinations has been successful. The objectives of the military, however, are among the clearest objectives in the field of education. A military task is to put the tread back on a tank or identify a friendly airplane's silhouette, not understand why the tonal system developed as it did.
27. If Arizona's integrated assessment program and Vermont's portfolio program only obtained reliability that ranged from .28 to .60, with math and English, we can have initial confidence that portfolios are not acceptable devices for high stakes assessment in music. (Don't ask me what a high stake is in music. Is it a high stake if students are prohibited from playing in a concert for not knowing the music?)
28. It takes little imagination to guess that the content of music portfolios might be quite different from one fourth grade class to another and perhaps the teacher's emphasis on the items within the portfolio might vary as well. A major task for us is to establish a shared interpretative framework. This will not be easy as two thirds of the students classified capable on the basis of a writing portfolio were not so classified on the basis of a direct writing prompt. Observations of laboratory work in science did not predict good performance on the simulation. In a science research project, it was found that students' performance was similar regardless of whether the assessment task was a hands-on problem or a paper and pencil facsimile of the problem. Thus, multiple tasks are needed.
29. Based on the New Standards Project of 1993 mathematics trial data, 9 to 25 items were required to achieve acceptable reliability and Shevelson concluded that at least 10 different topic areas of items may be needed to provide a dependable measure of a student's performance in one subject area. It is estimated that demonstrating competence on each item will take at least two class periods. Will we have assessment of a student's musical competencies without ignoring the difficulty of the task? It is likely that to take advantage of demonstrations, our instruction must change and even question whether teaching individuals or groups is our highest priority.
30. Other issues worthy of consideration include the fact that teachers provide more help to lower-ability students, exemptions are often made for special students, and the whole host of safety nets that teachers and administrators establish whenever value judgments form the basis for decision making. Social promotions and remedial classes are two examples. Vermont classroom teachers devoted 17 hours a month to finding portfolio tasks and evaluating the contents of the portfolios. Portfolios are a serious investment with costs of portfolios estimated from their use in advanced placement tests to be \$65 per subject as compared to \$2 to \$5 per subject for traditional measures. When the student creates his or her own response as in our creativity and improvising standards the complexity of assessing and scoring becomes even more challenging. It is critical that SRIG members conduct serious research on these compelling issues. Our

(continued on page 24)

THE EVOLUTION OF STATE ARTS ASSESSMENT: FROM SISYPHUS TO STONE SOUP

SCOTT C. SHULER, CONNECTICUT STATE
DEPARTMENT OF EDUCATION &
SELENA CONNEALY, COUNCIL OF CHIEF STATE
SCHOOL OFFICERS

The quantity and quality of effort around large-scale arts assessment have increased dramatically during the past few years. Among the factors that have contributed to this increase have been:

the general movement toward educational accountability in states;
the growing political awareness and assertiveness of supporters of arts education;
the development of National Standards in the Arts;
the National Assessment of Educational Progress in the Arts (NAEP); and
the State Collaborative on Assessment and Student Standards in the Arts (SCASS).

The movement among states toward test-based accountability has become almost universal. Alarmed by reports that American education is inferior to that of other industrialized countries, and increasingly skeptical of published tests that have allowed virtually every state to claim to be above average (sometimes referred to as the "Lake Wobegon Syndrome"), legislators and members of the public have called for more rigorous tests to determine how well their students are *really* doing. According to Education Week, "at least 32 states and 34 big-city districts now have accountability systems based, in part, on test scores. And the numbers are climbing fast."¹

The increased focus on measuring student learning as a means of motivating and monitoring school improvement has sometimes threatened and at other times buoyed the arts, depending on whether

the arts have been deemed "basic" enough to assess. It has become axiomatic in this new, assessment-driven climate that "what is tested is taught." Highly publicized state testing programs have had a particularly strong impact on local priorities as school administrators have tried to avoid the wrath of their communities by raising students' scores. Teachers are under pressure to teach "The Basics" — defined in practice as what is on the state tests— and to spend less time on "the other stuff" — defined as what is *not* on the tests. For the arts to be taken seriously in such a climate they must be identified as basic; to be treated as basic at the local level they must also be assessed, thereby holding districts or schools accountable for arts learning.

The debate over whether the arts are basic intensified during the Bush administration as two powerful groups of policymakers, the National Governors' Association and the U.S. Department of Education, developed "competing" sets of national goals for education, both of which neglected to mention the arts. After the election of Bill Clinton and his more arts-supportive administration, arts education organizations and their allies seized the opportunity to include the arts within the core curriculum outlined in the federal legislation known as *Goals 2000*, which was passed in 1994 and paved the way for standards-based educational reform. Their advocacy efforts showed a great deal more sophistication and unity than in the past, resulting in funding for the development of National Standards in the Arts² and for the concurrent development of the National Assessment of

Educational Progress (NAEP) in the Arts. When the latter assessment was administered in the Spring of 1997 it was the first arts NAEP since the 1970's and the first ever to address dance and theatre.

Although the National Standards and NAEP/Arts have been completed, the publicity campaign mounted by national arts education organizations has continued, gathering research and testimonials supportive of arts education and making sure that they reach the public through a variety of media. As a result, lawmakers and the public have become somewhat more supportive of including the arts in discussions of "basic" education. This has, in turn, supported the efforts of those who have advocated the inclusion of the arts in their state assessment systems.

Two major national projects have taken advantage of the existence of National Standards to advance the state of large-scale arts assessment. The teams developing the 1997 NAEP/Arts assessment had an opportunity to review and give input into the standards to ensure that they were clear and assessable. The content outline for the NAEP was then based on those standards. The State Collaborative on Assessment and Student Standards in the Arts (SCASS/Arts), a consortium of 19 states seeking to develop model arts assessment tasks, also used the National Standards as a consensus framework to describe the skills and understandings that should be assessed, thereby working around differences between individual state arts frameworks.

NAEP/Arts

NAEP/Arts has already contributed to the development of state assessments in a number of ways. First of all, the successful administration of the 1997 NAEP/Arts has proven that performance-based arts assessment can be conducted on a large-scale basis. Second, the development of NAEP/Arts greatly advanced the state of the art in arts assessment.

Useful NAEP/Arts products range from the innovative framework and specifica-

tions documents³ to the assessment tasks themselves. The artistic process model on which NAEP/Arts was based was proposed by Scott Shuler to the assessment design team initially as a way of finding common ground among the National Standards in the four arts disciplines for purposes of assessment, but has since proven to have important implications for standards-based instruction as well. As presented in the NAEP framework, the model suggests that students must learn to carry out the processes of creating, performing, and responding to each arts discipline (see figure 1). These artistic processes have provided the unifying foundation on which the SCASS/Arts project and individual states, such as Minnesota and Kentucky, have built their own assessments. The release of the NAEP items—tentatively planned for the fall of 1998—should provide models that will also be very useful to state assessment teams.

Perhaps even more important to the advancement of arts assessment is that the process of designing, administering, and scoring NAEP/Arts provided an opportunity for arts educators from across the country to develop their personal assessment expertise, while also generating products of potential use at the state and local levels. NAEP/Arts used an innovative approach to item development, gathering teams of artists and arts educators from SCASS/Arts participant states for training in task design, then relying on those teams to generate ideas for assessment tasks. Participants in the design and scoring of the NAEP/Arts assessment have carried that expertise back to their states, and have brought it to bear on SCASS/Arts and other state assessment initiatives. Some personnel who worked on NAEP/Arts have been employed by major education agencies such as WESTED, where they serve as advocates for including the arts in assessment initiatives; others have become professional assessment consultants who provide expertise to local and state education agencies.

SCASS Project

The goal of the SCASS project is to improve the quality of arts assessment in participating states. Since the project began in 1993, at one time or another 29 different states have been involved as partners. Each state education agency which participates in SCASS/Arts sends one content (arts expert) consultant and one assessment consultant to the group's quarterly work meetings. By collaborating to pool their limited personnel and resources, paying dues of only \$10,000 per year, member states have shared their expertise and have been able to accomplish much more than they could have individually.

Participating states have helped develop, pilot, and refine model assessment tasks linked to the National Standards in all four visual and performing art forms for students at grades 4, 8, and 12. Teachers in the SCASS/Arts states have piloted and suggested improvements in those SCASS tasks. Beginning in the summer of 1998, with funding from the National Endowment for the Arts, SCASS/Arts will conduct "trainer of trainer" institutes designed to help arts educators in each participating state develop assessment expertise. These educators will then return to their states to conduct inservice on assessment for other arts teachers.

One of the ancillary benefits of SCASS/Arts has been the collection of student work to illustrate national and state arts standards in the four arts disciplines. Such exemplary work will play an important role in preservice and inservice teacher training. Another benefit of SCASS has been its cultivation of a group of assessment consultants in state departments of education who are committed to, and advocate for, the inclusion of the arts in their states' assessment programs. Such internal advocates play an important role in furthering the cause of the arts in state agencies which often have only one arts specialist. Yet another benefit has been the opportunity SCASS meetings provide for arts consultants from participating states to communicate on an

ongoing basis, sharing information about current initiatives and projects. This has enabled many states to emulate innovations developed elsewhere and to avoid "reinventing the wheel" when solving common problems.

Statewide Arts Assessment: Scaling Sisyphus' Slippery Slope

At the individual state level, the ebb and flow of education reform initiatives have alternately fostered and undermined arts assessment. States tend to develop assessment of the core curriculum in a "classic" order of priorities that most educators can recite by rote:

English/language arts
math
science
social studies
other areas — including the arts— as
time and funding permit

As a result, virtually every state has implemented a testing system which includes periodic assessment in language arts and math. Unfortunately, however, few if any states have been able to sustain their education reform initiatives long enough to "get around to" the arts. States may be, as some have claimed, the "laboratories for education reform," but those labs tend to change ownership or close down before their experiments are finished. Governors run for higher office (or wind up in prison), legislatures change majority party, and state superintendents leave to take jobs with private foundations.

Other factors inherent in large-scale assessments tend to derail such initiatives. All statewide testing programs are expensive, and performance tasks are particularly expensive to administer, so legislatures have sometimes rebelled against providing or sustaining necessary funding. In some states there has been a backlash against the time devoted to statewide tests which some educators view as taking time away from instruction. States that have included performance tasks in their assessment systems have struggled

with the challenges of meeting traditional standards of reliability.⁴ All of these factors have caused some assessment-based reform initiatives to be scaled back before their arts component is in place.

The sequence of events in most states tends to be a variation on the following:

Initially, no plans for arts assessment.

Then a statewide education reform initiative is launched —by the state education agency, a group of business people, an ambitious legislature, or an “education governor”— which stresses accountability and assessment in all core subjects. These may take the form of assessment at various grade levels (Illinois), or perhaps a performance-based high school graduation assessment as an alternative to the traditional practice of counting credits (Minnesota, New York). Most of these initiatives are well-intentioned, but many are rather naive about issues of feasibility and/or cost.

At this point there tends to be a battle about whether the arts are included among the core subjects. The arts win a few and lose a few. (In Massachusetts, for example, arts advocacy efforts were successful in the legislature, but were derailed by a governor’s veto.)

If the arts win, then tentative plans are made to assess them, although usually after assessments are “on-line” in the traditional core subjects.

The state changes leadership, runs out of funds for assessment, or runs into a controversy that sidetracks the process (Kentucky).

Either statewide arts assessment plans are eliminated, arts assessment responsibilities are officially delegated to the local level, or the statewide assessment is limited to a relatively inexpensive multiple choice measure.

When considered from the perspective of each individual state, this process can seem quite frustrating. Arts educators interested in arts assessment sometimes feel like Sisyphus, the figure in Greek

mythology who offended the gods and was therefore doomed to roll a stone up a mountain, only to have it slip from his grasp each time just short of the summit. In state after state, arts supporters have painstakingly pushed the heavy stone of arts assessment up the steep slope of state priorities, only to see it fall away just short of implementation.

The good news, however, is that by the time a state’s arts assessment initiative changes direction, educators in that state have learned something about arts assessment which they then contribute to the general pool of expertise in arts assessment. The resources that the state has devoted to arts assessment have cultivated in-state interest and expertise (such as among teacher participants on statewide planning committees), and sometimes have yielded prototype tasks. As a result of many such experiences, and the contributions of national projects such as NAEP and SCASS, the over-all, aggregate direction of arts education *nationwide* has been one of progress. Imagine a line graph in which each state’s arts assessment efforts rise and fall, but the total (nationwide) trend is upward.

Survey Data and Trends

Our survey of states’ plans for arts assessment shows that a number of states have specific plans to conduct statewide arts assessment. Data gathered via that survey are summarized in Table 3. The aforementioned ebb and flow of reform initiatives means that it is likely that the information in the table is already out of date — somewhere a state has either announced or canceled plans for arts assessment. The news over-all, however, is surprisingly encouraging, with 12 states either administering or developing arts assessments.

State arts assessment plans vary widely in purpose and design. However, the data reveal a few commonalities and provide some basis for discussing trends.

Several states which are participants in SCASS mention the development of state

frameworks or standards as an important step toward conducting assessment.

Of the 12 states reporting either an operational arts assessment or, more commonly, plans to administer such an assessment, most are planning assessment in all four visual and performing arts.

A few states have attempted to address the need for breadth and depth expressed in the National Standards. The Standards proposed that each student develop a broad background in all of the arts K-8, then develop depth in at least one art form through study at the high school level. State assessments in Minnesota and Illinois allow students at the high school level to select one art form in which to demonstrate their expertise.

All of the states which either have arts assessment in place or are far enough along to have item specifications are using multiple choice for at least a portion of the assessment. The multiple choice format offers a cost-efficient way to measure some learning, but clearly cannot address many important artistic behaviors such as the ability to create or perform art works. As outlined in the NAEP/Arts specifications document, a comprehensive approach to assessing arts achievement should include a variety of assessment formats, including performance tasks. Given the alternative of having no arts assessment at all, however, arts educators in several states have made the tough decision to support a 100% multiple-choice assessment.

No state has sustained regular performance-based assessment in the arts at the statewide level. However, several states that are planning arts assessment either hope to administer performance assessment statewide or will mandate that such assessment occur at the local level.

Kentucky is an interesting case study. After an earlier, flawed attempt at performance-based assessment in the arts, crippled in part by the requirement that the assessment be interdisciplinary, the state is working its way back toward per-

formance assessment in stages. Kentucky's department of education has created a new assessment framework, directly tied to national standards and built on the artistic process model developed originally for the NAEP/Arts assessment. The framework calls for assessment to be developed in three phases:

- Phase 1: Totally pencil-and-paper: both the test items (prompts) and students' responses are limited to what can be expressed on paper.
- Phase 2: Multimedia stimulus, pencil-and-paper response: test items can involve sounds and video as well as still visuals, but student responses are limited to paper.
- Phase 3: Multimedia stimulus, performance (multimedia) response: both test items and student responses can be multimedia with student responses recorded on audio or videotape for off-site scoring.

It will be interesting to see whether Kentucky, which has thus far sustained its unusually expensive, "high stakes" assessment system, will be able to stay the course. Assessment results in Kentucky have an enormous impact on individual schools, influencing their level of state funding, their teachers' salaries and tenure, and the job security of their principals. Such a system has, predictably, whipped up a great deal of resistance which has in turn created political pressure for change. In fact, as this article was being written, Kentucky's state senate had voted to enact a number of changes.⁵

Predictions and Suggestions

There is reason to doubt whether any state will be able to afford annual performance-based arts assessments adminis-

tered to every student, even if limited to only one grade level. Arts assessments are generally not "high stakes" (i.e., do not tend to carry strong rewards or consequences). It is difficult to imagine parents allowing low arts achievement scores to prevent their students from graduating, although precisely such an assessment system is planned in New York and Minnesota. It is also difficult to imagine a school's arts assessment scores affecting its funding or the job security of its teachers, although that is part of the plan in Kentucky.

A more likely function for statewide arts assessments will be to provide an accountability tool to monitor the success of local districts or schools. A practical design would be one in which students are sampled to derive data about learning either statewide, to compare a state's achievement to the nation as a whole (as in Minnesota's and Connecticut's readministrations of NAEP music items circa 1980), or to measure performance at the district or school building level (as in Kentucky). A recent survey of attitudes among education managers and policymakers conducted by the National Center for Education Statistics, entitled *NAEP's Constituents: What Do They Want?*, reveals that 56% were either "probably interested" or "definitely interested" in participating in a so-called State NAEP in conjunction with the next NAEP/Arts assessment, currently planned for 2007. A State NAEP oversamples participating states, thereby permitting state policymakers to compare their students' achievement in a subject to that of students nationwide.⁶ The results of this survey further substantiate the increasing support among state agency personnel for arts assessment.

Arts assessments for the purpose of local accountability would not have to be administered annually, because student achievement would not change significantly during the few months between the date when one year's scores were finally scored and published and the date of the next test administration. It might therefore be possible to administer statewide

arts assessments on a rotating basis, such as:

one year dance, the next music, then theatre, then the visual arts; or one year arts, then social studies, then foreign language, then physical education; or one year all of the arts at grade 4, the next at grade 8, then at grade 10 or 11.

Entrusting local districts with administering and scoring state-developed assessment tasks, or even developing their own tasks, reduces the likelihood of deriving reliable enough data to make comparisons of student performance between districts or schools. Such a solution does, however, encourage districts to develop local expertise in assessment. That is, in fact, the proposed strategy in Minnesota, where state-developed tasks are intended to serve as models, but districts are free to develop and score their own performance tasks to determine whether students have met graduation requirements in the arts.

Other Signs of Progress

Beyond those states with formal plans for arts assessment, the good news is that the level of expertise in arts assessment is improving in many parts of the country. Assessment experts at Educational Testing Service and in various state education agencies have commented that they are intrigued by the challenges and possibilities of assessing arts achievement.

The Educational Testing Service, which won the contract to design and score NAEP/Arts, is very excited about what it has learned through this innovative assessment. ETS personnel have commented repeatedly during the development, administration, and scoring of the 1997 NAEP/Arts that they have learned an enormous amount in the process both about the arts and about performance assessment. They have had to develop new methods of carrying out virtually every phase of the assessment process. For example, the 1997 NAEP/Arts moved

away from stand-alone items to integrated sets of items in which various formats (multiple choice, short answer, performance tasks) were combined around a particular work of art or concept. A particular music task-set might, for example, ask students first to answer a few relatively easy multiple choice questions about a listening excerpt, then to provide a longer "constructed response" by comparing that excerpt to another listening example, and finally to perform or create a variation on one of the excerpts to which they have listened. Scoring such integrated item sets, meaningfully aggregating (totaling) those scores, and developing appropriate reporting strategies have challenged traditionally trained ETS statisticians to take advantage of the latest cutting edge statistical and reporting strategies. This newly developed "in house" expertise has led the organization to consider expanding its involvement in arts assessment, such as by developing additional Advanced Placement arts exams.

Now that the SCASS project has begun circulating model tasks among arts educators in participating states, and states such as New York and Minnesota have begun to make their model arts assessment tasks available on web sites (see addresses at end of article), access to quality assessment strategies is becoming more universal. Other states that have no plans for uniform statewide arts assessment are, nevertheless, encouraging local professional development on the topic of assessment, toward developing local models.

In Connecticut, for example, progress in assessment is occurring on several levels. The state has awarded Goals 2000 funding to a number of local districts for the purpose of developing arts assessment tasks and using those tasks to collect student work in relation to its new state arts standards which are themselves linked to the National Standards. The Connecticut Music Educators Association (CMEA) is piloting and refining assessment tools for elementary schools to measure, and to gain recognition for, the

extent to which they are achieving the state standards through their general music programs. Through research and adjudicator training seminars, CMEA is engaging performing ensemble teachers in the process of developing and refining analytic scoring rubrics for its annual all-state adjudication process.

Connecticut is also piloting portfolio-based teacher assessment, as part of its Beginning Educator Support and Training (BEST) program, which beginning art and music teachers must pass to renew their teaching licenses. The portfolio, itself a model of performance-based assessment, requires beginning teachers to teach their students to carry out the three artistic processes, assess their students' achievement, and reflect on both their own and their students' work. Series of beginning teacher seminars are designed to develop teachers' assessment expertise. Training of the experienced arts educators who develop, refine, and score the portfolios, and the mentor and cooperating teachers who work with beginning teachers, will eventually extend the culture of assessment and self-reflection statewide. [Web site for more information: <http://www.state.ct.us/sde/brta/index.htm>]

How to support statewide arts assessment in your state

The following are strategies that will increase the chances of statewide arts assessment in your state:

Find funding for your state to participate in SCASS. Although the \$10,000 annual participation fee should ideally be provided by the state education agency such as through its Goals 2000 initiatives, there are many other alternatives. In some states, the arts council or a professional arts education organization has provided the necessary funding; private foundation support is also a possible avenue.

Publicize the results of the 1997 NAEP/Arts and use it to foster public awareness of arts education and of the fact that the arts can be assessed and are important

enough to assess. Cooperate to administer NAEP/Arts items statewide to permit comparison to national achievement.

Lobby for inclusion of arts in the state's core curriculum and assessment programs.

Garner arts council or foundation "challenge" support to elicit state department of education funding for statewide assessment development.

Perhaps most important of all, develop a "culture of assessment" among arts educators in your state. Having arts educators who are better prepared for classroom assessment will lead to the following developments: more support existing at the grass roots level for assessment; more expertise available for statewide assessment design and implementation committees; and policy makers who are more aware of the feasibility and desirability of arts assessment.

Toward this end, make use of local, state, and federal funding for workshops to train teachers and artists about assessment (and about teaching and learning in general). For example, such professional development is an appropriate use of federal Eisenhower and Goals 2000 dollars.

Conclusion

The primary purpose of assessment is to improve teaching and learning. In states with strong test-based accountability systems, however, inclusion of the arts in statewide testing is also essential for political reasons because subjects that are tested are treated as "basic." Fortunately, well-designed statewide assessment efforts bring other important benefits such as fostering a culture of assessment and reflection at the classroom level.

The road to state assessment in the arts has been long and often frustrating, typically characterized by hard work, hopeful developments, then disappointment. Experienced arts educators are, however, resilient folk who refuse to be deterred

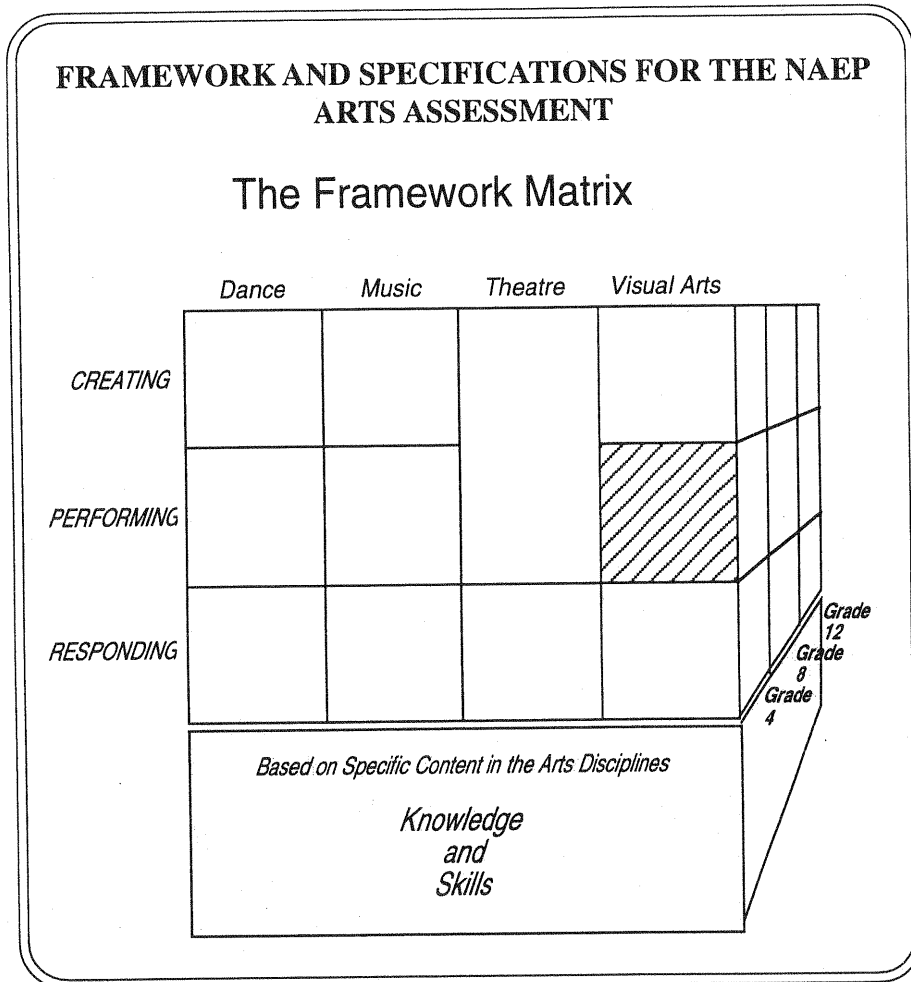


fig.1

by frequent setbacks. It is therefore finally possible to see success on the horizon. States have drawn on the model of NAEP/Arts and shared their expertise, through projects such as SCASS/Arts, to make the development process easier and less expensive. As a result of such collaboration and hard work we are finally leaving the Sisyphus metaphor behind, choosing instead to make "Stone Soup." Let us hope that the effort brings us closer to the high standards we have set for ourselves and our students.

Resources for further information

Information about Minnesota's assess-

ment tasks and plans can be found at <http://children.state.mn.us/grad/grathom.htm>

Information about New Yorks assessment tasks and plans can be found at <http://www.nysed.gov>

Information about NAEP can be found at <http://nces.ed.gov/naep/index.html>

Information about SCASS/Arts can be obtained by writing to Selena Connealy c/o CCSSO, One Massachusetts Ave., NW Suite 700 Washington, DC 20001-

1431, by calling her at (202) 336-7076 or e-mailing her at selenac@ccsso.org

ENDNOTES:

- 1 Olson, Lynn. *The Push for Accountability Gathers Steam*. Education Week, vol. 17 no. 22, 2/11/98, pp. 1-13.
- 2 Music Educators National Conference. *National Standards for Arts Education: What Every Young American Should Know and Be Able to Do in the Arts*. Reston, VA: MENC, 1994.
- 3 National Assessment Governing

(continued on page 24)

MUSIC ASSESSMENT FRAMEWORK

ARTS PROCESSES FOR MUSIC

table 1

<p>CREATING - <i>When improvising, composing, or arranging music, students:</i></p> <ul style="list-style-type: none"> - APPLY HISTORICAL, CULTURAL, AND AESTHETIC UNDERSTANDING BY CREATING STYLISTICALLY APPROPRIATE ALTERATIONS, VARIATIONS, AND IMPROVISATIONS; - USE STANDARD AND/ OR NON-STANDARD NOTATION TO EXPRESS ORIGINAL IDEAS; - EVALUATE, REFINE, AND REVISE SUCCESSIVE VERSIONS OF ORIGINAL WORK; - DEMONSTRATE SKILL AND EXPRESSIVENESS IN THE CHOICE AND USE OF MUSICAL ELEMENTS; AND - PRESENT THE CREATED WORK FOR OTHERS.
<p>PERFORMING - <i>When singing or playing music with musical instruments, students:</i></p> <ul style="list-style-type: none"> - SELECT APPROPRIATE REPERTOIRE; - APPLY SKILL BY PERFORMING WITH TECHNICAL ACCURACY; - DEVELOP AN APPROPRIATE AND EXPRESSIVE INTERPRETATION BY APPLYING UNDERSTANDING OF STRUCTURE AND CULTURAL AND HISTORICAL CONTEXTS OF MUSIC; - READ MUSICAL NOTATION ACCURATELY; - EVALUATE, REFINE, AND REVISE THE PERFORMANCE; AND - PRESENT THE PERFORMANCE FOR OTHERS.
<p>RESPONDING - <i>When perceiving, analyzing, interpreting, critiquing, and judging music, students:</i></p> <ul style="list-style-type: none"> - SELECT REPERTOIRE FOR LISTENING; - ANALYZE THE ELEMENTS AND STRUCTURE OF MUSIC; - COMPARE AND CONTRAST VARIOUS MUSICAL STYLES; - IDENTIFY FORMAL AND EXPRESSIVE QUALITIES THAT DISTINGUISH A PARTICULAR STYLE OF MUSIC; - PLACE MUSIC WITHIN ITS CULTURAL AND HISTORICAL CONTEXT; - MAKE CRITICAL JUDGMENTS ABOUT TECHNICAL AND EXPRESSIVE QUALITIES OF MUSICAL PERFORMANCES AND COMPOSITIONS; AND - USE MOVEMENT OR WORDS TO INTERPRET AND DESCRIBE PERSONAL RESPONSES TO MUSIC.

table 2

BASED ON SPECIFIC CONTENT IN MUSIC	
<u>KNOWLEDGE</u>	<u>SKILLS</u>
<i>APPLYING KNOWLEDGE OF:</i>	<i>APPLYING COGNITIVE, AFFECTIVE, AND MOTOR SKILLS INCLUDING:</i>
CONTEXT:	PERCEPTUAL
-PERSONAL	INTELLECTUAL/ REFLECTIVE
-SOCIAL	EXPRESSIVE
-CULTURAL	TECHNICAL
-HISTORICAL	
AESTHETICS	
FORM AND STRUCTURE	
PROCESSES	

LARGE SCALE ARTS ASSESSMENT IN THE STATES

MARCH 1998

STATE	DISCIPLINES	GRADES	PURPOSE(S)	SAMPLE	STATUS	ITEM TYPE (%)	HOW ADMINISTERED	COMMENTS
IL	D, M, T, V	4, 7, 11	2, 4	MATRIX SAMPLE OF ALL STUDENTS	IN PLACE	MULTIPLE CH. (100)	COMPONENT OF GENERAL ASSESSMENT	BASED ON STATE MANDATE. ARTS ARE ONE OF THE SIX AREAS OF LEARNING.
KY	D, M, T, V, L	5, 8, 11	1, 2, 3, 4, 5	MATRIX SAMPLE	IN PLACE	MULTIPLE CH. (80) EXTENDED RESP. (20)	STAND ALONE ASSESSMENT	EXTENDED RESPONSE SCORES IMPACT A SCHOOL'S ACCOUNTABILITY INDEX. WILL MOVE TO MORE EXTENDED RESPONSE IN FUTURE (UP TO 90%).
MA	V, L	4, 8, 10	1, 2, 3, 4, 5	ALL STUDENTS	IN PLACE	MULTIPLE CH. (50) SHORT ANSWER (25) EXT. RSP. (25)	COMPONENT OF GENERAL ASSESSMENT	IN LIT. PORTION OF ENGLISH/LANGUAGE ARTS EXAM AND ART HISTORY QUESTION IN HISTORY EXAM. OTHER DISCIPLINES MAY EVENTUALLY BE ASSESSED AT LOCAL LEVELS (BASED ON STATE DEVELOPED STANDARDS)
MID	N/A	N/A	N/A	N/A	IN PLANNING	N/A	N/A	TASK FORCE WORKING TOWARD BUILDING ARTS ASSESSMENT INTO MSPAP (STATE'S SCHOOL ACCOUNTABILITY MEASURE). AT BEGINNING STAGES OF DEVELOPMENT.
ME	D, M, T, V	4, 8, 10	2, 5	MATRIX SAMPLE	IN PLACE	SELECTED RESP. (30) PERFORMAN-CE (70)	COMPONENT OF GENERAL ASSESSMENT	THE NEWLY ADOPTED STANDARDS WILL MAKE THE VISUAL AND PERFORMING ARTS A "BASIC" PART OF MAINE'S LEARNING RESULTS STANDARDS AND ASSESSMENT PROGRAM.
MN	D, M, T, V	9-12	1, 2, 6	ALL STUDENTS	PROPOSED	PERFORMAN-CE (100)	ADMINISTERED AT CLASSROOM LEVEL	MINNESOTA BOARD OF ED HAS ADOPTED GRADUATION EXAMINATIONS IN READING, MATH, AND WRITING. HAS PROPOSED TO ADOPT 10 ADDITIONAL AREAS OF LEARNING TO WHICH INCLUDE THE ARTS. STUDENTS WOULD BE REQUIRED TO CREATE AND/OR PERFORM IN THE ARTS AND WOULD BE ASSESSED AT THE CLASSROOM LEVEL USING A STATE-DEVELOPED RUBRIC.
MO	D, M, T, V	3 GRADE LEVELS	2, 5	TBA	VOLUNTA-RY 2000, REQUIRED 2001	MULTIPLE CH. EXTENDED RESP. (%TBA)	STAND ALONE ASSESSMENT	ARTS WERE NOT DESIGNATED AS A CORE SUBJECT AREA BY THE LEGISLATURE, BUT THE STATE BOARD OF EDUCATION HAS DESIGNATED THEM AS PART OF THE CORE AND ASSESSMENT IS BEING FUNDED AS A RESULT OF THAT DESIGNATION.
NY	D, M, T, V	9-12	1, 2, 3, 4, 5	ALL STUDENTS	2001	MULT. CH. (25) PERE. (75)	STAND ALONE ASSESSMENT	PLANS FOR HIGH SCHOOL GRADUATION REQUIREMENT.
OK	M, V	5, 8, 11	2, 5	ALL STUDENTS	FIELD TEST 97-98, IN PLACE 98-99	MULTIPLE CH. (100)	N/A	PART OF PRIORITY ACADEMIC STUDENT SKILLS.
PA	D, M, T, V	4, 7, 10	2, 5	TBA	PILOTING	MULT. CH. (30) PERE. (70)	TBA	DEVELOPED AS PART OF FEDERAL GRANT. WILL EVENTUALLY BE USED AS MODEL FOR CLASSROOM LEVEL ASSESSMENT.
UT	V	1-6	5	VOLUNTA-RY AT SCHOOL LEVEL	IN PLACE	PERFORMAN-CE (100)	STAND ALONE ASSESSMENT	PART OF CORE CURRICULUM
WA	D, M, T, V	7-12	N/A	TBA	2004	N/A	STAND ALONE ASSESSMENT	PART OF EDUCATION REFORM

TABLE 3 (KEY: D=Dance; M=Music; T=Theater; V=Visual; L=Literature)

PORTFOLIOS: ASSESSMENT, PEDAGOGY, OR STAFF DEVELOPMENT?

**LYLE DAVIDSON,
NEW ENGLAND CONSERVATORY OF MUSIC
AND LINCOLN CENTER INSTITUTE**

As work moves ahead on the National Standards, it is appropriate to consider what role or roles portfolios might play in that work. I want to consider the current situation in portfolio use, specifically, their use in schools, their use in teaching, and their use in assessment.

Let's think about why, in spite of such promise in the beginning, they are not used more widely; think about their strengths and weaknesses; and try to advance some ideas that may make them more useful in schools and work with students.

Review of portfolio practice

Thirteen years ago, the arts group at Harvard Project Zero began an attempt to develop a systematic approach to assessing the arts (visual arts, music, and creative writing). We decided to look at arts assessment practices already in place which naturally led us to consider portfolios.

We did not understand what a portfolio might look like in music and it took us three years to establish our working principles. The first year we spent simply observing classes and talking with the teachers about particularly interesting instances of teaching and learning. The second year we proposed some approaches to developing evidence of the learning we observed and began to document strategies on a pilot basis. The third year we worked by successive approximation (revision on demand) and began to codify both the process of developing portfolios and the products a portfolio would contain. The final product, the Arts PROPEL portfolio, contains domain projects, questionnaires, interviews, and

journals (Davidson, Myford, et al, 1992).

The use of portfolios was well established in the visual arts. More recently, English and writing classes have been using portfolios, or as they are more typically known, writing folders. For 15 or more years, the teaching and assessment of writing has been under review. As a result of that work (which used portfolios) our understanding of the nature of writing has been transformed.

Over the past ten years, portfolios have also become a brand name, a trade word. They have become linked to writing programs, to critical thinking programs, and to writing across the curriculum efforts. Also, the term, "portfolio" carries implications of alternative and/or authentic assessment. That association conveys that knowledge is being used in rich contexts and may involve the ongoing review of work that is possible with formative assessment models.

Portfolios are not restricted to a single level of schooling.

Although currently developed for use in elementary and secondary levels, portfolios have penetrated tertiary levels of education in some places. Portfolios are a core component of students' education at Alverno college; they have replaced proficiency examinations in the writing programs at NYU at Stony Brook and University of Minnesota; they play a key role in the solfege program at New England Conservatory, and Northwestern requires students in music education program at the masters level to maintain a portfolio.

Portfolios are used in public school staff development programs organized by the

Bernstein Center, the Lincoln Center Institute, and the Department of Education in Wisconsin.

Finally, because portfolios have been found to be valuable in school change efforts, staff development workshops in portfolio use are now part of nearly every summer program for in-service teachers, at least in elementary education.

To sum up, the language of schooling has changed a great deal over the last 10 years. The term "portfolio" as part of a broader authentic assessment effort has become a common term. And portfolio use is not restricted to one level of education. They are used across a range of school programs, elementary, middle, high school, tertiary, and graduate schools and in a number of other contexts as well.

A problem

The showcases we hear about provide rich examples of how portfolios can be used, and they do suggest what is possible when portfolios are included in instructional programs. But, exciting though they appear, they may be false positives. Showcases do not illustrate the extent to which portfolios are actually used.

If you visit a large number of schools, you realize that very few classrooms actually use portfolios. Portfolios, in spite of their value, are not widely used.

Why are they so scarce?

I see four reasons why portfolios are so hard to find, and why they are not adopted by every teacher, principal, and district: 1) Portfolios are not well defined; 2) they are not connected to the dominant testing paradigm; 3) they may not match the educational philosophy that guides the teaching in a given classroom or school, and finally, where they are linked to school change, there is usually too little support for the goals and process of the change.

1) Definition of portfolios

First, portfolios are not well defined. Portfolios are sometimes difficult to see because they can be one thing in one context and something very different in another. Also, they can hide as authentic or performance assessment. Given this lack of clarity, it is worth the effort to make some distinctions among the uses of portfolios, and to create some distance between assessment and portfolios.

From my perspective, there are three kinds of portfolios (Davidson, 1993). Portfolios can reflect best work, or consist of a collection of work, or they can be more closely allied to educational practices. Let's consider these for a moment.

a) *Best works portfolio*: A best works portfolio is familiar as a typical artist's portfolio. In music, the audition represents this type. The best works portfolio contains final products. It makes relatively efficient measurement possible, it is quick and focused. However, there is no evidence of thinking beyond the single work or the single achievement; the working style is hidden and the range of expertise is limited to the sample of work chosen.

b) *Collections of work portfolio*: The collections of work portfolio is a place where work is kept. The artists sketchbook (or Roger Clemens' journal in which he documents the characteristics of every player he comes up against) are examples of this. The collections of work portfolio contains a range of work. Because of this, it tends to reflect the curriculum (or the characteristics of the individual). However, without a set of orienting guidelines except association, it is not designed to show systematic evidence of learning. It provides no guidelines for viewing the contents, and no framework for interpreting what it contains.

c) *Process portfolio or Learning portfolio*: The Process/Learning Portfolio exists as a record of one's learning. It contains the records of the romance of engagement, the mastery of facts and specifics that follows, and the growing abil-

ity to interpret and apply general principles of a discipline (its fundamental ideas and relationships, its structure, its organizing principles). The process portfolio also provides guidelines for ongoing review of the contents as well as rubrics for interpreting the work. The Arts PROPEL music portfolios are of this type.

This third type, the *process or learning portfolio* is what I consider a portfolio to be. It shows the process of learning, the use of information and concepts, and provides evidence of growth. The other portfolio types, *best works* and *collection of work*, are more exhibits—at least in my mind. Of course, exhibits are also useful—but that is another paper.

2) Mismatch with assessment practice

The second reason portfolios are not common may be that there is a mismatch between portfolios and the needs of assessment. They are not part of the testing paradigm. Most standardized tests still depend on identification, short answer questions, and solving problems that are decontextualized and simplified. Such tests are good for looking at knowledge acquisition.

Like standardized tests, portfolios can provide evidence of knowledge acquisition, and they are far more useful tools for looking at knowledge use. They are useful for documenting the problem-solving process and showing the thinking that leads to the solution. Knowledge-use problems tend to be rich and not as sharply defined or constrained as knowledge-acquisition problems. They are more like problems encountered in everyday work, the kind of problem that engineers encounter in the field, or musicians find in the practice room.

Portfolios are useful because the documents they contain can reveal what the student did to understand and interpret situation or problem, especially when the thinking required involves comparing, sorting, organizing, and making sense of a situation.

3) Mismatch with classroom practice

The third reason portfolios may not be used in schools is that they may not match the educational philosophy or practices. Given the emphasis on the *how* as well as the *what* of knowledge and understanding, portfolios may not easily fit the predominant theory of instruction in classrooms where the focus is on knowledge acquisition, child-centered programs, or on developing social skills.

A survey of teachers working with the Lincoln Center Institute suggests they base their classroom practices on four different orientations, a child-centered approach, a social/interpersonal approach, a more constructivist discipline focused approach, and a traditional approach to teaching.

Forty percent (40%) identified their practices as child-centered, "I let the children's interests and what excites them guide me." or "I try to find ways to reach and use each child's learning style." Portfolios, when there are any in these classrooms, tend to consist of samples of an individual student's work. These typically hang on the classroom wall. I suppose such portfolios could be thought of as *best works portfolios* for a single class.

Thirty-two percent (32%) described their classrooms as places where hands on or experiential learning took place. They spoke about a constructivist viewpoint; identified coaching as important ("... where the teacher lays the foundation and the children construct their own inquiries."); and talked about classrooms that involve children in planning, inquiry, decision making, reflection and (hopefully) taking ownership of learning. This is the group of teachers who are most interested in portfolios. These are the classrooms in which portfolios show knowledge use, the *learning portfolio* type. In any case, in these portfolios one finds more emphasis on developing the ability to think, organize, communicate, and evaluate work.

Twenty-five percent (25%) of the teach-

ers stressed the importance of positive interaction and cooperation among children. They spoke about using different approaches to teaching and about trying new things to help all students participate in classes. They felt that the well-rounded class used the arts whenever possible to facilitate this. These classrooms tend to have exhibits of examples of students' best work hanging on the walls.

Only three percent (3%) of the teachers described themselves as traditionalist when teaching. They described their classrooms as "very practical and basic," and they speak about the importance of mastering of basic facts and skills as well as the importance of teaching students how to study and work. These teachers use writing journals that are closely related to teaching writing. The focus in these classrooms is on knowledge acquisition.

The results of this survey suggests that schools may not be as homogenous as they may appear. The principals of the schools in which the surveys were carried out are seeking ways of bring a more experiential model of teaching and learning into their building. But, as we know, classroom practices tend to reflect the experiences and propensities of the individual teacher, not the vision of the principal.

Also, the survey indicates that more teachers are interested in aspects of learning that best fit traditional methods of assessment, i.e., in which the portfolio process is not necessary. Knowledge acquisition tends to be the focus of those classes. Only 32% of the teachers declared that their classroom practices included place for the students to participate in planning, inquiry, decision making, and reflection on their work.

4) Lack of support

Finally, there is relatively little support for staff development. Shortage of funds, time, space, materials, combine with lack of thorough introduction and follow-

through: In general, districts provide very little support to help teachers change their classroom practices.

It seems that many districts and teachers have unrealistic expectations as they search for new solutions to problems. Teachers want tools that easily fit what they currently do; while at the district level, staff development tends to be considered as an off-the-shelf commodity that can be purchased in one day doses. The assumption appears to be that once the staff development pill is consumed, teachers can transform practices built up over years of work, change deeply held values, and make easy transfer from the simplified context of the workshop to the complexity of the classroom. However well intentioned the district and state efforts may be, teachers rarely get sufficient support for change.

Furthermore, portfolios take time to develop. Typically, teachers who use portfolio practices in their classrooms are at first dismayed by the amount of time they take away from traditional instruction. Those who stay with portfolio practice, and discover ways to accommodate the initial time commitment, begin to change their classroom practices as they make portfolio work a more integrated part of their work.

Given these problems, the lack of definition, the mismatch with tests, the mismatch of educational philosophy, and the lack of support, why not give it all up? Portfolio practice is still being developed. Problems, such as these I have suggested, are only forces that are helping to shape our understanding of the opportunities and limitations of portfolios. As we gain experience with portfolios, we may be able to help solve such problems—or at least reframe them in more productive ways.

Framework for portfolios

How do we overcome these problems? Before offering solutions, let's consider what portfolios do really well and what do they not do as well. Let's develop a

framework for considering portfolios.

What do portfolios do really well? Portfolios can be powerful tools for documenting learning. The portfolio is flexible. It can contain various kinds of documents, tests, essays, notes and sketches, tapes of rehearsal, and drafts and completed compositions—and the results of standardized tests, if you choose.

Portfolios have very high validity (*Editor's note: Validity is interpreted differently with portfolio assessment. In traditional evaluation an assessment must be reliable in order to be valid*). The experience in Arts PROPEL, the Wisconsin initiative, and our experience at the New England Conservatory make this very clear. When repeated samples of work are taken, the process of learning and attendant changes that occur becomes a salient feature.

Portfolios are good tools for raising standards within the classroom. By engaging students in developing the criteria for self-evaluation, making critique of their own work over a long period part of the class work, and by making the range of work taking place in a classroom more visible, quality begins to become everyone's criteria for assessment; students begin to see their work as a source of pride.

Because portfolios can document learning so well, they can be used to show the community what kind of work goes on in school (as at Shutesbury Elementary, in Shutesbury, MA).

Finally, portfolios can be strong tools for creating a climate of reflection in a school. They provide occasions for teachers to gather around student work and explore their own values and goals. This can be a novel agenda in schools where the meetings are typically called because of low test scores.

What do portfolios not do as well? Portfolios have very low reliability. This makes them difficult to use for district level assessment, or any situation that

requires measurement and quantitative data. Such work requires tremendous support from the district and state. The experience of assessment in England should make us pause before we use portfolios in this way.

Portfolios are not effective without a great deal of support. Typically, insufficient time is designated for the discussions required to develop rubrics, score examples, and modify classroom practices.

Portfolios are not quick to use. Effective use of portfolios requires time during the school day for conversations between teachers and students, and to be really effective between groups of teachers. School schedules do not provide the time needed for this.

Portfolios may work less well in an environment in which the cultural transmission model is the predominant approach to learning. At the New England Conservatory, those teachers who place their emphasis on accuracy in performance tend to not see portfolios as useful. Portfolios thrive best (if not require) an environment in which there is an emphasis on participation and active learning, where there is a constructivist emphasis and an environment in which knowledge use is important.

Making portfolios useful

Now, how can we make the advantages of portfolios more evident and therefore, portfolios more useful. Three questions need to be asked as portfolios are being considered: First, define why portfolios are going to be used; second, decide what type of portfolio is going to best meet the goals and objectives; and finally, determine how they are going to be implemented and what is going to happen once they are in place. These suggestions may sound obvious and simple to carry out, but in schools, where action and activity tend to drown out reflection, these suggestions are often not considered. Conversations with principals, who want to bring portfolio practice to their schools, demonstrate how hard it is to stop the

action long enough to be able to think about how to address these steps.

1) Define why the portfolio is going to be useful. What is their purpose

Why are they being used? Is the intention to use portfolios to replace standardized tests? as a tool for assessing student work? as a tool assessing teachers classroom work? to improve teacher practices? or are portfolios going to be a tool in a school change effort that requires more reflective practice?

Again, the use of portfolios for high stakes assessment at the district level is probably not a good idea. Without major changes in the culture of tests and measurement, portfolios are not going to meet the criteria for assessment and evaluation at the district and state level. The experience of using portfolios statewide in Vermont and Kentucky shows us how hard it is to gain reliability and to make them useful instruments for making policy. Samples from portfolios can be used to help the public or policy makers interpret and show the context of the results of standardized tests.

Portfolios can be used for several different purposes:

- Creating a classroom environment that honors reflective practice. The Arts PROPEL materials are rich with examples of this.

- Improving teachers' classroom practices. At PS 191 in Manhattan, four times during the year over a two year period, teachers gathered samples of their students' work completed within the Lincoln Center Institute Focus School program and discussed them with one another. They also studied works of art as texts during this time. The principal says that this staff development program played the most important role in getting the school off the SURR (School Under Registry Re-

view) list. Had their test scores not improved, the school was going to be closed.

- A tool in a school change effort. At PS 116 in Manhattan, the principal wants to create stronger and more visible connections among the parents, teachers, and students. Portfolios are being used as part of that process.

2) Define what the portfolio is and how it fits into the educational environment.

Arts PROPEL provides rich examples of what a process portfolio can be (Davidson, Myford, et al. (1992). Let me remind you of the characteristics of those portfolios. Arts PROPEL portfolios:

- are based on processes that artists use (making, discriminating, and thinking about work), thereby providing a variety of windows or perspectives on musical products and processes;

- contain multiple samples of work collected over time and long term projects that are engaging to teachers and students;

- feature a variety of types of work (not simply performances, but what students thought about those performances, and how they used what they learned in other contexts);

- use a variety of strategies for sampling work: journals, domain projects, periodic surveys of work in progress

- provide frameworks and specific criteria for reviewing the contents of the portfolio on an ongoing basis, revealing how work changes over time.

3) Think about how the portfolio will begin and how the outcomes will be followed up.

It is important that sufficient time be given to the development of the practices that support a portfolio culture. This may mean a good deal of time outside of the regular school day to ensure concentration on the task. Before beginning the process it is important to think about what will happen to the portfolios themselves; where they will be stored; whether or not they will be cumulated; and what needs to be taken into consideration when they be begin to be effective.

It is important to realize that using portfolios can change our understanding of what we do. This can be an opportunity or a threat. Consider what happened in the teaching of writing as a result of the portfolio process. The change from thinking that "writing is writing" to unpacking the skills and strategies for writing led to a new understanding of how those skills and strategies change depending on the purpose and audience of the writing. We now realize that writing is an extended process that takes place over a long time and draws on many different approaches to thinking and expression during that process (Camp & Levine, 1991).

In this time of National Standards for music, it is important to recall that the changes in understanding of writing created a split between methods of teaching and methods of assessing what is written. We may have to make a similar shift in music, a shift from saying that "music is music" or "music is performance" to advancing a view that is far more dimensionalized, if we develop a portfolio culture and make use of it for questioning the nature of musical experience as it develops.

Conclusion

At Project Zero, the Arts PROPEL portfolio effort began as an alternative to standardized, high stakes testing—at least in Pittsburgh¹. Thirteen years later, I have a far more articulated view of portfolios. Now I see that

- Portfolios are a useful tool for

improving teachers' classroom practices and pedagogy;

- Portfolios are valuable tools for initiating and guiding school change efforts;
- Finally and maybe most importantly, portfolios are strong tools for creating a climate of reflection in a school.

Are portfolios about assessment then? If the portfolios contain only work samples, journals, and self-reflections, they are probably not useful for high stakes state-wide assessment—at least not without enormous resources and time devoted to designing a holistic scoring process that is reliable. On the other hand, a portfolio can certainly include class tests, quizzes, final examinations and standardized tests. In either case, a portfolio process that allows students to play a formative role in developing the framework for review certainly can raise the quality of work that takes place in the classroom.

Are portfolios about pedagogy? They certainly can have an impact on classroom practice. Portfolios become occasions for dialogue between teachers and students. These conversations are grounded in the students' work showing the teacher how knowledge is acquired and how it is used.

Are portfolios about staff development? In my experience, this has certainly been the case.

Bibliography

Anson, C. M., Brown, R. L. (1991) Large-scale portfolio assessment: Ideological sensitivity and institutional change, in (Pat Belanoff & Marcia Dickson, Eds) *Portfolios: Process and Product*. Portsmouth, NH: Boynton/Cook Publishers. 248-279.

Camp, R. & Levine, D. L. (1991). Port-

folios evolving: Background and variations in sixth- through twelfth-grade classrooms, in (Pat Belanoff & Marcia Dickson, Eds) *Portfolios: Process and Product*. Portsmouth, NH: Boynton/Cook Publishers. 194-205.

Davidson, L. (1993). Portfolio assessment and Exhibits: Moving from recall to reflective understanding. *Holistic Education Review*, Spring, 6.1 45-51.

Davidson, L., Myford, C., Plasket, D., Scripp, L., Swinton, S., Torff, B. & Waanders, J. (1992) *Arts PROPEL: A Handbook for Music*. Cambridge, MA: Harvard Project Zero.

Dollase, R. H. (1996). The Vermont experiment in state-mandated portfolio program approval. *Journal of Teacher Education*, March-April. 47, No. 2.

Scott, P. (1991). Step by step: The development in British schools of assessment by portfolio, in (Pat Belanoff & Marcia Dickson, Eds) *Portfolios: Process and Product*. Portsmouth, NH: Boynton/Cook Publishers. 80-92.

Travalen, Fern. (1995). *Vermont Arts Assessment Project: Focusing on the nature of artistic practice in learning*. Montpelier, VT: Vermont Council on the Arts.

¹ Editors' note: *No standardized high stake test could be identified as having been used in music in Pittsburgh. The reference is to language arts. We were also unable to determine any present use of music portfolios in Pittsburgh for assessment purposes.*

AUTHENTICALLY ASSESSING THE NATIONAL CONTENT STANDARDS IN MUSIC: THE VALUE OF PORTFOLIO ASSESSMENT AND PAPER-AND-PENCIL TESTS

PATRICIA E. SINK, SCHOOL OF MUSIC
UNIVERSITY OF NORTH CAROLINA -
GREENSBORO

Assessing outcomes of music education provides numerous opportunities. Boyle and Radocy (1987) maintain that measuring and evaluating music learning helps to establish music as a viable curricular area holding students and teachers accountable for standards of achievement. Colwell (1979) explains that testing and measuring music behaviors improves music instruction by providing powerful feedback to both teachers and students. As a result of assessment, teachers are provided specific evidence of student progress and of accomplishments resulting from music instruction. When applied appropriately, assessment functions as a powerful learning tool for students. Assessment, therefore, should be an integral part of music education providing effective tools for developing music life skills and knowledge.

What are music life skills and knowledge? In 1994, the National Committee for Standards in the Arts announced the first voluntary national content standards in arts education for grades K-12 in the United States (Music Educators National Conference, 1994). The MENC Task Force for National Standards in the Arts, chaired by Paul Lehman, identified content standards in music including skills, knowledge and understandings from prekindergarten through twelfth grade. Ultimately, the national content standards in music describe the musically educated adult upon completion of fourteen grades of education (i.e., preK-12). Since the development of the content standards in music and the 123 associated behavioral objectives, several concerns have been identified. Particularly important to the

content of this Newsletter, the content standards and associated behavioral objectives emphasize the need for reliable, valid, and appropriate techniques for assessing students' music learning. A variety of techniques may be used to accomplish valid and reliable assessment of student learning in music including objective paper-and-pencil tests and techniques often associated with portfolio assessment. The music education profession is relatively familiar with using paper-and-pencil tests to measure students' attainment of behavioral objectives. Portfolio assessment, however, is a relatively new technique, particularly in terms of its effective and efficient application in classrooms and rehearsal rooms.

What is portfolio assessment? In the current newsletter, Lyle Davidson provides a detailed and accurate description of portfolio assessment, including its strengths and weaknesses. Unfortunately, some music educators, however, seem to believe that portfolio assessment results in "authentic assessment" to the exclusion of objective, paper-and-pencil tests.

Valid assessment of musical learning is not possible if the teacher is limited to multiple-choice items or to paper-and-pencil tests. "Authentic assessment" — . . . refers to assessment techniques that require demonstration of the specific behavior or skill sought in the objective rather than a paper-and-pencil representation of it — . . . (MENC, 1994, p.5).

Some music educators may misunderstand such descriptions of assessment as meaning that objective paper-and-pencil

tests are not authentic measures of music learning. Of course, the preceding quote is taken out of context, and the MENC (1994) publication, *The School Music Program: A New Vision*, does not explicitly support the exclusion of traditional forms of assessment. Indeed, the MENC Task Force for National Standards in the Arts stressed the need for "authentic assessment." Upon examination of the national content standards in music and the associated behavioral objectives, however, a reader should readily perceive that authentic assessment includes both objective and subjective information and representations of music learning obtained via traditional and alternative forms assessment.

As previously indicated, a variety of techniques are needed to assess students' attainment of the behavioral objectives associated with the nine content standards in music for grades K-12. Below is a paraphrased and abbreviated listing of the nine national content standards in music identified by the MENC Task Force (MENC, 1994, pp. 1-2).

1. Singing
2. Performing
3. Improvising
4. Composing
5. Reading and Notating
6. Listening, Analyzing and Describing
7. Evaluating
8. Understanding relationships between music and other disciplines
9. Understanding music history and culture

What are techniques that may be used to authentically assess students' attainment of the behavioral objectives associated with the nine content standards in music? Numerous behaviors associated with content standards 1 through 4 may be assessed authentically via various observational techniques often associated with portfolio assessment. Some of these techniques include:

1. Checklists and anecdotal reports describing verbal and nonverbal music responses;

2. Systematic observations describing students' time-on-task, number of occurrences of music behaviors, and students' tendencies to participate in music experiences;
3. Ratings and verbal descriptions of the quality of students' music products (i.e., performances and compositions) relative to degrees of accuracy, originality and involvement, and relative to the development and evolution of student products via a variety of processes; and
4. Students' self-assessments and descriptions of their development of musical life skills.

Numerous behaviors associated with content standards 5 through 9, in part, may be assessed via objective, paper-and-pencil tests using an assortment of types of test items including multiple-choice, short answer, fill-in-the-blank, matching, true-false with explanation, and essay items. Results of applying the aforementioned assessment techniques producing objective and subjective information should be a compilation of students' accomplishments of the content standards in music for grades K-12. Results of broadly defined portfolio assessment which includes information attained via objective, paper-and-pencil tests and via formal and informal observation techniques should result in a portfolio of students' abilities to function as adults using music knowledge, skills and understandings throughout life.

What are the benefits of portfolio assessment? Establishing standards and methods of portfolio assessment should provide music educators valid and reliable techniques via tests and observational systems to document students' learning and development as music performers, composers, and consumers. Additionally, portfolio assessment, as well as the national content standards in music, have focused music educators' attentions on the need to develop and use valid and reliable assessment techniques. Applications of valid and reliable assessment

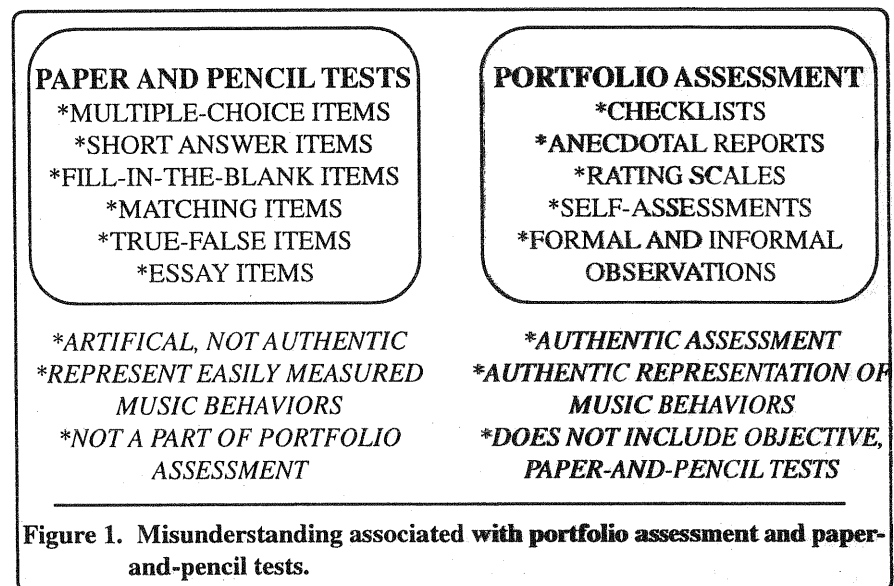
techniques should help music educators to:

1. facilitate and improve music instruction;
2. determine the effects of instruction on music learning;
3. facilitate students' music learning and development;
4. provide necessary and powerful feedback to students and teachers about music teaching and learning;
5. guide students toward developing music life skills and knowledge; and
6. verify music as a curricularly viable subject holding teachers and students responsible for attaining standards of excellence.

What are some problems associated with

Are there behaviors associated with the nine national content standards in music that may be most effectively and efficiently measured via objective, paper-and-pencil tests? Excluding the prekindergarten content standards and behavioral objectives, there are 109 behaviors associated with the nine content standards identified by the MENC Task Force (MENC, 1994, pp. 13-26). Of these behaviors, 41 behaviors (38%) may be most effectively and efficiently measured via objective, paper-and-pencil tests. Table 1 (page 20) includes selected K-12 content standards in music and associated behaviors that may be validly measured via paper-and-pencil tests.

Of the 109 behaviors associated with the nine national content standards in music for grades K-12, 68 behaviors (62%) may be most effectively assessed via tech-



portfolio assessment? Among some music educators during the past decade, the concept of portfolio assessment, viewed as *the* "authentic assessment," seems to have inspired an attitude that objective, paper-and-pencil tests are artificial and are not authentic representations of music learning. Figure 1 illustrates this misunderstanding that portfolio assessment should exclude traditional forms of assessment.

niques often associated with portfolio assessment (see Figure 1 below).

Based upon an analysis of the nine national content standards in music and their associated behaviors, it is apparent that music educators need a variety of techniques to authentically assess students' attainment of the content standards and the associated behavioral objectives, including paper-and-pencil tests. When music educators misunderstand "authen-

TABLE 1
SELECTED K-12 CONTENT STANDARDS AND BEHAVIORS APPROPRIATE FOR PAPER-AND-PENCIL TESTS

- *K-4: Content Standard 5 (Reading and Notating) (p. 15):**
- c. identify symbols and traditional [music] terms...
 - d. use standard symbols to notate [music elements]...
- Content Standard 6 (Listening, Analyzing, and Describing) (p.16):**
- a. identify simple music forms when presented aurally
 - d. identify the sounds of a variety of instruments,...as well as voices
- *5-8: Content Standard 8 (Understanding Relationships...) (p. 20):**
- a. compare two or more arts...
- Content Standard 9 (Understanding Music History and Culture)(p.20):**
- b. classify [musical works] by genre and style...and explain...
- *9-12 Content Standard 6 (Listening, Analyzing, and Describing) (p.24):**
- a. identify/explain compositional devices [which create unity and variety]

(MENC, 1994, PP. 13-26)

tic assessment" and conclude that portfolio assessment excludes the use of objective, paper-and-pencil tests, they eliminate a powerful means for assessing approximately one-third (38%) of the behaviors associated with the national content standards in music. Determining if music educators are developing intelligent performers, composers and consumers of music, in part, is dependent upon applying valid and reliable paper-and-pencil tests.

Another problem with portfolio assessment is that applying effective observation and subjective reporting systems is time-consuming, perhaps more time-consuming than developing valid and reliable paper-and-pencil tests. Related to this problem is that assessing via observation and subjective reporting systems seems to work best in individualized and small-group instructional settings. Typically, music education in public schools is provided via large-group instruction with class and ensemble sizes ranging from 25 to 150 students. Music educators and researchers need to identify strategies for efficiently and effectively applying observation and subjective report-

ing systems in large-groups as a part of instruction.

How do music educators overcome some of the problems of portfolio assessment?

Training music educators to effectively use and efficiently apply a variety of assessment techniques in music instructional settings is essential to assess music learning authentically. As a result of training, music educators should begin to value assessment and measurement as essential instructional tools. Additionally, music educators need to avoid the tendency to define traditional forms of assessment as artificial; rather, objective, paper-and-pencil tests are essential to valid and reliable portfolio assessment. The Music Educators National Conference should establish teams or task forces of educators and researchers to develop the various techniques essential to a valid and reliable portfolio assessment system. Three task forces may be sufficient to develop an effective and efficient system.

Assessment Task Force 1: develop and test systematic observation techniques and subjective reporting systems for assessing process and product behaviors

associated with content standards 1 through 4 and with selected behaviors associated with content standards 5 through 9.

Assessment Task Force 2: develop and test paper-and-pencil tests to assess auditory and auditory-visual perception skills and music knowledge identified by behaviors associated with content standards 5 through 9.

Assessment Task Force 3: develop and establish a system for maintaining composite and continuing records of students' music learning from prekindergarten through grade 12. This task force should develop music development and learning charts and forms for formally compiling the objective and subjective information attained via the assessment methods developed and tested by Task Forces 1 and 2.

Finally, measurement specialists and active researchers of MENC need to conduct research to begin answering assessment questions identified by the MENC Research Task Force, chaired by

Carolynn Lindeman. The Research Task Force compiled the document entitled, *A Research Agenda for Music Education: Thinking Ahead* (MENC, 1998). Within the document, the assessment questions listed below are identified.

1. What are valid forms of assessment in the arts and how can they be used to improve student achievement in music?
2. What are the expected musical and other outcomes of a good music education and how can they be documented?
3. What are some strategies for incorporating ongoing assessment into ensemble rehearsals and general music classes? How can individual learning, as well as group learning, be determined within context?
4. How can technology be used to facilitate record keeping and assessment? MENC, 1998, p. 9)

“Music assessment should be based upon specific objectives that identify clearly what students should know and be able to do” (MENC, 1996). Ensuring that assessment of music learning is linked realistically with music life skills and knowledge is the responsibility of music educators. Whether assessing via paper-and-pencil tests, observation techniques, rating scales, checklists, anecdotal reporting systems, or self-assessment techniques, the goal of assessment efforts is to effectively guide and direct students toward independent musicianship as adults. Any assessment system may be artificial or may be authentic — the authenticity of assessment, in part, depends upon its effective and efficient application and use; thus, continuing debates about assessment authenticity or artificiality seems pointless. A focus of the music education profession is to ensure that all students have equal opportunities to function independently as musically educated adults. One means of doing this is by providing detailed and realistic feed-

back resulting from valid and reliable assessment. Such feedback and assessment may be used to promote music growth and development and to facilitate the best music instruction, and thereby, provide each student opportunities to participate in music throughout life as a performer, composer, and/or consumer.

References

Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. New York: Schirmer.

Colwell, R. (1979). *Silver Burdett music competency tests - Books 1-6:*

Teacher's guide. New Jersey: Silver Burdett.

MENC (1996). *Performance standards for music: Strategies and benchmarks for assessing progress toward the national standards, grades preK-12*. Reston, Virginia: Author.

MENC (1998). *A research agenda for music education: Thinking ahead*. Reston, Virginia: Author.

MENC (1994). *The school music program: A new vision*. Reston, Virginia: Author.

**MUSIC EVALUATION TODAY:
OF PENDULUMS, TUG-OF-WAR, TAILS WAGGING
DOGS, COMFORT ZONES, AND THE
RELEVANCE OR IRRELEVANCE OF
AUTHENTIC ASSESSMENT**

**BRIAN LEAVELL, SELWYN SCHOOL
DENTON, TEXAS**

As the snooze alarm went off the other morning, my wife, who is normally swift to hammer silence into the foul mechanism, paused for several seconds to listen to the random radio snippet: “It doesn’t matter if they’re three, fifty-three, Anglo or Hispanic, I can teach anyone to read with phonics”. Of course, lately the probability has risen dramatically that the topic at any given moment on a radio talk show would involve someone advocating phonics. As California outlaws whole language, other states begin to advocate phonics as well - [and the pendulum swings back]. Another bipolar trend seems to involve some educators beginning to favor total immersion in English, while others hold fast to the more gradual approach using bilingual education [swing, swing, pendulum swing]. Old school researchers get bent out of shape over operational definitions of ‘partici-

pant-observer’ and many other qualitative givens, while qualitative folk question objectivity and other quantitative givens[tug, tug, we win you lose]. As some sing the praises of portfolio assessment, others say, “that’s another oxymoron to add to my list”. It can be seen time and again - the progressivists and the essentialists and/or perennialists doing an educational tug-of-war. And, just as in the game, it is unacceptable to many to leave the little red flag in the middle of the playing field - even if it’s the best thing for students. The middle ground between qualitative and quantitative assessment is a healthy place for assessment because both have unique strengths. A balance between group and individual assessment is likewise advisable.

But we need to back up just a bit. Before we place ourselves somewhere within the

swing of the pendulum between quantitative and qualitative assessment [portfolios as authentic assessment here for brevity], we need to look at the authenticity of our teaching practices and learning outcomes. First, let's take a look at the musical experiences we provide in school. The performance experiences we provide to middle school and high school students are what I call 'institutional music making'. Why do we have such a huge drop-out rate after high school? I believe it's due in part to the fact that there are no performance vehicles in society that reflect what we do inside the institution. Take away the institutions, educational, military and religious, and you will find nary a concert band or a choir. There are community bands and I'm told by the director every time I go to a concert that these rarities are growing. But these bands still have nothing to do with the child's musical education outside the institution. Orchestras exist in society, and we need to continue to do our best to keep them alive [without subversive education]. But what are we doing with these dated ensembles in school? In band, we've constructed and utilized 'objective' tools to help kids determine what instruments are best for them. In reality all we've done is [knowingly or unknowingly] use an invalid and unreliable assessment tool to direct the child toward a musical instrument that will allow us to have balanced instrumentation in a performance vehicle that hasn't been popular in society in 60 to 70 years. The group is determining the child, with little or no balance of performance vehicles that will allow the child to determine the group. Said another way, the child is being assimilated into an institutional musical reality instead of having their societal musical reality accommodated - a pendulum that should be stuck in the middle instead of on one side. Further, in order to determine whether or not the director is doing an adequate job, we've created criterion-referenced tools utilized by judges to check groups against some standard and the judges often end up comparing bands to one another in a much more subjective manner than can best be described as a bad example of norm-ref-

erenced assessment. Because judges don't want the standard set too high too soon, performing first, second, or third on the first day is the kiss of death for those who want the top rating. Listening to directors in state after state worry about who performs first before whom assures me that these things are a reality. I've even had a judge tell a friend that if she had played two bands later, he would have given her band the top rating. Already we are out of balance and our assessment pendulum starts to swing toward a preponderance of group assessment. Solo and ensemble festival helps to balance this pendulum but the same assessment hazards are still there - as is the same hazard in group evaluation of having students play repertoire that was created exclusively for this institutional music reality. What's worse is that we are no better than the classroom teachers who are forced by their administrators to use study halls and even class time to 'teach to the test'. [Administrators and teachers being evaluated by the students' test scores is a validity problem in itself, as is band directors being evaluated exclusively by their students' performances, but that's another topic] When we spend most or all of our classroom time getting ready for competition or evaluation, the evaluation is determining the curriculum, pollution of the students' score on the assessment tool is highly likely, and the tail is wagging the dog.

The instrument makers have constructed 'objective' tests to determine musical aptitude when in reality these tests are little more than recruiting tools. And another pendulum starts to swing toward focusing on performance assessment and little attention paid to assessment in non-performance areas of music education. And, what about those non-performers? If they have a music class at all beyond elementary school, it's usually 'general music' [whatever that is] and the students are relegated to being 'informed listeners'. We're going to 'enlighten' them by showing them how to have an 'aesthetic experience' with Mozart and other good music, and with blind faith believe that I applaud Frances Rauscher for standing

up during the MENC conference and reminding us to interpret her findings with several cautions, such as remembering the difference between causation and correlation.

But we are improving their temporal-spatial acuity simultaneously. When was the last time you listened to Matchbox 20, Sublime, or Puff Daddy, if for no other reason, to inform yourself of your middle schooler's or high schooler's musical reality? Remember, the root word of ignorance is ignore. If you ignore popular music, you are making consciously uninformed judgments and generalizations about vocalists in their chest voices, repetitious musical forms, and other common complaints, you are creating a musical experience that is out of balance. We should be getting these non-performers re-involved via recording technology, music marketing, music merchandising, and other musically related activities instead of being content with 13 to 20% participation in performance ensembles. Since our penchant is focusing on the musical product and not the process [a pendulum that's stuck and needs some oil], how about a program where live performances and CD's are the final products and kids who don't perform are responsible for production, recording, press releases, marketing and merchandising the product? When we decide that we're going to pay attention to non-performers in a way that doesn't make them second-class musical citizens, then we should worry about trying to assess their work.

Elementary music education, instrumental and general, is predominated by adaptations of 100 year old methods from other countries [Kodaly, Orff, Dalcroze, and Suzuki]. When students sing or play American folk music [on authentic instruments], they are carrying on an important authentic tradition [even if the venue is not authentic]. When we utilize pentatonics in improvisation on Orff instruments [an insitutional instrument] so students don't experience dissonance, are we doing them a favor or a disservice? And, is this practice authentic to an edu-

cational or a societal reality? When we sing pentatonic repertoire because we've been told that children cannot sing half steps in tune, are we relying on years of systematic or unsystematic observation to guide us? Am I suggesting we exclusively incorporate authentic practice so we can use authentic assessment? No, I am suggesting that we have a balance between societal and institutional music experiences and the respective assessment tools that best evaluate these experiences.

So, here we are. Determining evaluation tools for performance and trying to balance the pendulum that swings favorably toward group assessment. If we choose portfolios because we want authentic assessment then we should not use them because we are not utilizing authentic practice because validity involves measuring what you purport to measure. I believe one purpose of portfolios in music evaluation might be to balance the process/ product and group/ individual pendulum. As an instructional tool, portfolios can add a focus to the process of music making as they are a form of comprehensive musicianship [more depth and breadth]. If students include a tape of their best performance, they could also include a journal entry or a conference with the teacher utilizing a 'think-aloud protocol' in order to gain insight into the cognitive processes students incorporate when they play. This information, while not assessment per se, could be used to facilitate and assess curricular and/ or lesson objectives which focus on cognitive processing during performance. If getting students to be critical of their individual performances is an instructional focus, then portfolios can be used to measure this self-criticism. Students could be required to rate their own individual performances on the tape they submit. This rating could be compared to the teacher's rating to see areas where the student's musical self-concept is lacking. But again, one must look at what the portfolio can ASSESS before using it as an assessment tool and it can only assess what it can instructionally facilitate. But caution should be heeded if using a port-

folio for instruction or assessment. One of the problems I have encountered with implementing portfolios in the middle school classroom involves goals and goal setting. Middle school students have some difficulty with long range goals. Therefore, short-range goals [one to two weeks] should precede, and act as scaffolding to long-range goals. Goal setting should initially be done by the teacher, giving the students freedom with option underneath that goal. An example of a portfolio goal would be to have students learn more about their instrument beyond performing on it. Options underneath this goal could include exploration into the instrument's acoustical properties, consummate performers, the history of the instrument, or anything else that gets neglected as a result of an exclusively performance-oriented education. Later, once goal setting has been modeled and you get your middle school students accustomed to thinking about more long range goals, then they can begin to set goals themselves.

It appears that standards or competencies will play a key role in educational reform and they seem to be a good way to combine the elements of qualitative and quantitative assessment. Several things need to be considered, however. First, with respect to the competencies themselves, they should involve concepts and skills derived from the repertoire. Performance of the repertoire itself should not be a competency. Second, they should be consistent with some national, state or local curriculum framework. Third, there should be some glossary of terms which, through their definition, helps to reduce ambiguity. Fourth, the competencies should be agreed upon by an expert panel. Performance standards or competencies must have rubrics which distinguish the various levels of performance from one another. First the rubrics should go beyond changing one or two words that distinguish the categories; and, if terms like all, most, or some are used, they should be defined. Second, rubrics should be specific to a certain competency, not one-size-fits-all. Third, the rubrics should include an unacceptable category so the

lowest acceptable category doesn't become a dumping ground for every performance that doesn't qualify for the higher categories. Last and like the competencies themselves, rubrics should be agreed upon by an expert panel. An example of an application of these guidelines is applied below.

Competency: Guitarists should physically locate and verbally identify note names on the G and D strings from the first through the twelfth fret.

Rubrics:

Advanced - Immediately (within less than two seconds) upon hearing the fret number named, move hand directly to the fret and name the note.

Proficient - Diatonic scale or dotted frets are used as landmarks to move hand to proper fret and note is named within 2 - 5 seconds.

Basic - Chromatic scale or every fret is used to count up [or down] to the correct fret and the note is named within 5 - 7 seconds.

Unacceptable - Assistance is required with finding correct string, or utilizing chromatic scale in order to determine correct response.

In conclusion, I believe the entire profession should be required to take a course in dialectic logic. For at least four months of our lives, we will be forced to take a thesis, its antithesis and argue to some middle ground. Far too often we [knowingly or unknowingly, intelligently or unintelligently] incorporate propositional logic and argue to our own myopic conclusions. And, these conclusions support practices which are comfortable to the teaching professional not the novice or group of novices s/he teaches. It is in these teacher comfort zones that we pose the greatest potential threat to individualized instruction and student assessment. If we teach only in our comfort zone, the students' comfort zones must be assimilated. Further, it is in these as-

Music Evaluation Today

Leavell

(continued from previous page)

assessment and/or research comfort zones, that we may be letting the assessment determine the curriculum and the methodology determine the research. If we choose research methods or assessment tools that only lie within our comfort zone, we clearly are letting the tail wag the dog.

Introduction

Colwell

(continued from page 4)

present research base is very small and with little activity by music evaluators during the past decade.

31. AP studio art portfolios were most valuable as a framework for evidence about skills and knowledge.

32. Vermont principals and other users see beneficial effects on schools in curriculum, instruction, and attitudes.

Much of the confusion occurs because of a failure to distinguish between student and program evaluation. For evaluators these are interesting times.

**State Arts Assessment
Shuler and Connealy***(continued from page 10)*

Board. *1996 Arts Education Assessment Framework. Pre-Publication Edition and Arts Education Assessment and Exercise Specifications for the 1996 National Assessment of Educational Progress (NAEP)*. Washington, DC: March 1994. [Available by calling The Council of Chief State School Officers at (202) 336-7076.]

4 Koretz, Daniel. *Vermont Portfolio*

Study: Don't Overstate the Problems. *Education Week*, 1/13/93.

5 Lawton, Millicent. *Ky. Senate Panel Approves State Test Overhaul.* *Education Week*, vol. 17 no. 23, 2/18/98, p. 18.

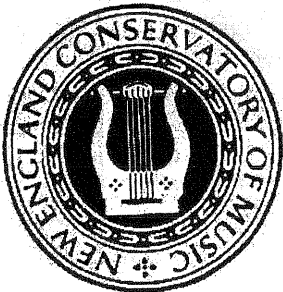
6 National Center for Education Statistics. *NAEP's Constituents: What Do They Want?* Washington, DC: U. S. Department of Education, Office of Educational Research and Improvement, January 1998, p. 31. [Available by calling (202) 219-1690.]

ANNOUNCEMENT

**INCOMING CHAIR AND
NEWSLETTER EDITOR:
DR. THOMAS GOOLSBY,
GEORGIA STATE
UNIVERSITY,
ATLANTA, GA, 30303**

**New England Conservatory
290 Huntington Avenue
Boston, MA 02115**

*Address Correction
Requested*



Attn: Richard Colwell and Greg Gatien

Non-Profit Org.
US Postage
PAID
Permit No. 57238
Boston, MA 02115