In This Issue:

Evaluation
Authors

Judy Dawson
State Department of Illinois

Keith Swanwick
Professor of Music Education
University of London,
Institute of Education

Carol Myford
Center For Performance Assessment
Educational Testing Service
Princeton , New Jersey

Editors
Richard Colwell, Jeff Roberts
New England Conservatory of Music
Boston, MA

Assessment: Some Trends and Cautions Based on One State's Experiences
Judy Dawson

Arts assessment—the process of collecting valid, reliable information about student learning in the arts—has become a nationwide concern. Many states and other entities such as schools, districts, colleges and universities, and professional organizations have begun to develop or provide assistance in arts assessment. At the national level, multiple efforts have focused on arts assessment:

professional organizations developed content and achievement standards that *"identify what our children must know and be able to do"* (Consortium of National Arts Education Associations, 1994, p. 6) and that can provide a basis for student assessment as well as program evaluation (p. 15),

national legislation (Goals 2000: Educate America Act) included the arts as a basic content area and encourages arts assessment,

the arts will be included in the National Assessment of Educational Progress (NAEP) assessment of 1997,

arts educators from throughout the country worked together to develop specifications for that assessment,

consortia of arts educators and state department of education consultants from 15 states collaborated to develop arts assessment tools for the NAEP assessment,

most of those consultants—who represent arts content and assessment work with consultants from other states in a related project (State Collaborative on Assessments and Student Standards (SCASS) Arts Education Consortium) through which they are developing arts assessment tools for large-scale and small-scale use,

several conferences (organized by professional arts education associations, the American Council for the Arts, and the Getty Center for Education in the Arts) have included multiple sessions on arts assessment, and

one of the five goals adopted by the National Art Education Association focuses on assessment. (The goal, "establishing criteria and means for assessment," is quoted in documents such as National Art Education Association, 1995.)

These activities provide evidence that leaders in arts education have become increasingly supportive of arts assessment. Grass roots support from practicing arts educators has also grown. Fewer and fewer educators challenge or resist assessment. Increasingly more seek resources to help them develop it. Many educators are convinced that instruction is likely to be more effective if they systematically collect and analyze information about student learning. They also appear to be convinced that the increased emphasis on the arts can help them obtain resources to improve and expand arts programs. They know that, in return, they are expected to show that the resources are helping students learn more in the arts.

This article will review and reflect upon arts assessment development activities in Illinois. That effort began after the Illinois legislature, in 1985, passed an educational reform bill that, among other things, named the fine arts (dance, drama/theatre, music, and visual arts) as one of six basic content areas and required state and local goals and assessment of all six areas. Initially, arts assessment was scheduled to include all students in four specified grades and to begin in 1992. Because the state department of education, known as the Illinois State Board of Education (ISBE),

would be assessing all students (approximately 100,000 in each of four grades) in six content areas, testing efficiency was critical. All tests would be machine scorable (except in writing, where essays would be used). It quickly became obvious that, since the state arts assessment would be multiple-choice, it would include only four of the five state fine arts goals—because one goal clearly referred to creation and performance in the arts and could not be assessed validly with a paper-and-pencil assessment.

Multiple-Choice Assessment

ISBE's first major priority was to develop a collection of assessment items for use in the state assessments. The process included several steps and would require at least 3-5 years to:

> —identify or write assessment items,
> —review the items,
> —pilot test the items with students,
> —review the results, and
> —begin the cycle again by revising items as indicated through pilot testing and by writing new items for the collection.

Arts educators from throughout the state have been involved in every stage—primarily to ensure that the items validly represent the state arts goals. Some became members of a fine arts advisory committee established to make recommendations about the content and format of the state assessment. Many helped as item writers or reviewers.

The process began by assembling arts tests from other states, NAEP assessments in the 1970's, and several private sources (at least two Illinois school districts and a commercial bank of assessment items). Then, Illinois arts educators reviewed the tests and identified test items they considered appropriate for assessing Illinois goals. The items were then pilot tested with students. Since that initial year, all items have been written by Illinois arts educators.

The process of developing assessment items has been enlightening. We've learned:

—Using music notation with items (for example, requiring students to examine notation as they answer questions about the expressive qualities of music) appears to enhance the validity of items, but such questions may be invalid or unfair for students who cannot be expected to know how to read notation. Music notation has an important function in testing, but must be used carefully.

—Items that appear to be most valid for assessing some arts learning because they require students to answer questions about how to use arts principles rather than simply about what those principles mean are often problematic. Artists can use arts principles effectively in many ways. Arts educators are often reluctant to identify a particular response as "correct." Also, distractors (incorrect answer choices) that do not compete as correct responses are so obviously wrong that very few students select them. Thus, the items do not function well in

multiple-choice tests.

—It is much easier to develop multiple-choice items that assess factual knowledge than higher-level thinking. It is easier to write items that focus on the role of the arts in history and culture (e.g., about major composers or forms of music) or on processes and tools required to perform in the arts (e.g., musical symbols and instruments) than on sensory elements, organizational principles, and expressive qualities. It is especially difficult to write good higher-level items that do not require prerequisite knowledge which cannot be assumed.

—The active involvement of arts educators in assessment development is critical. Their content expertise helps ensure that questions validly represent the arts. Without their high standards, the state arts assessment might have become just another assessment in another content area. Instead, the state assessment includes only content that can be assessed validly with a multiple-choice test. Test booklets include music notation, colored images of major visual artworks, and photographs of dancers. Videotaped assessments and thematic exercises (see below) are being developed. Beneficial side effects include participants' provision of assistance to schools, coordination of and presentations at conferences and workshops, development of arts assessment programs for schools, and general support for arts assessment.

—Developing assessments that are of high quality requires persistence and constant vigilance. Relatively few assessments are completely satisfactory the first time they are tried out with students. Virtually everything has to be revised at least once and tried out again.

At this point, Illinois has accumulated a basic collection of several thousand multiple-choice items. However, we plan to supplement those items with alternative assessment approaches. Development of two types (videotaped assessment and thematic exercises) has begun. As they become available, we propose incorporating them into the state assessment. Fortunately, a legislative change specifying that the state shall assess only samples of students in the arts gives us the addditional flexibility needed to use alternative assessment approaches with some students.

Videotaped   Assessment

At the beginning of the state assessment development process, arts educators recommended that arts assessment should include requiring students to look at/listen to actual artworks or performances rather than simply answer written questions about the arts. Consequently, images of major visual artworks were used with questions and included in test booklets. However, performances in the other three arts areas required sound and/or movement. Staff decided to develop videotapes that included excerpts from arts performances and questions about them. (This approach is similar to the audiotaped assessment commonly used in music but adds the video component—primarily because it is essential in dance and drama/theatre and we wanted to use the same approach in all three performance arts. Interestingly, music educators who initially suspected that the video would distract from the audio eventually became convinced that the video's advantages—e.g., enabling students to watch musicians play instruments and see the role of the conductor—more than outweighed its disadvantages.)

At this point, only one video—in music—has been completed. Its development was facilitated by the state department of education's participation in the production of an arts videotape series ("Arts Alive") which gave us the right to copy excerpts from that series. Unfortunately, the quality of many of the performances was limited. Regardless, a development committee identified excerpts to use and wrote questions for them. The tape was produced, pilot tested with students, revised, and piloted again. At that point, committee members decided to release it as an illustration of the videotaped assessment approach.

Since then, the committee has identified a videodisc series, obtained permission to use it, and written questions for excerpts. We are now attempting to identify other performances to include on a tape that represents more than the classical music on the videodisc series as well as the ethnic and cultural diversity of the state's residents..

We are also in the process of developing videotaped assessments in dance and drama/theatre. In all three arts areas, the major obstacle has been to identify existing tapes of performances that can be copied with-

out violating copyright laws. Obtaining permission to copy videotaped performances is more complicated than obtaining permission to copy written material. Videotaped performances often have multiple copyright holders (for example, video distributors, composers, choreographers, lyricists, arrangers). Also, royalty agreements with performers may influence copyright. In the search for usable videotaped performances, progress was significantly hampered by several false leads: commercial videos for which permission to copy was denied—despite the fact that several were produced by small or public ventures; a video distribution company official who granted permission to copy videos without having the right to do so; dance videos using copyrighted music; locally produced videos with performers whose dated hair and clothing styles would have distracted many teen-aged students; locally produced tapes with poor camera work or without original tapes from which good copies could be made; and, video producers without established procedures for granting copyright permission.

An alternative is to videotape local performances. However, it is still necessary to ensure that no copyrighted materials (such as music notation or arrangements) are used, that written permission is obtained from everyone who will appear on the videotape, and that the type and quality of the videotape will be suitable for making multiple copies.

Also, though, local educators can use videotaped assessment without producing new videotapes and thus avoid copyright problems. Educators can use existing videotapes—from a school's collection or rented from a local video rental store. They can identify excerpts from and write questions for the videotapes. They can use index numbers to identify the location of the excerpts, show the excerpts, and give students the test questions.

Many types of questions can be used in videotaped assessment. Essay questions—which would be scored systematically using a scoring rubric—can be especially useful for asking students to analyze, critique, or compare performances. Performance assessment can be used for students to demonstrate that they can perform in a style similar to performances they saw and heard. Forced-choice assessment (multiple-choice, true/false, or matching) can be used to assess whether students recognized how, for example, sensory elements of the arts were used in a performance.

Thematic Exercises

An assessment approach that was recommended for—and will be used in—the NAEP arts assessment requires students to complete an exercise that resembles a mini-project and that may include several types of assessment (but are often predominantly performance-based or constructed response). For example, students might play an unfamiliar piece of music, listen to a recording of their performance, and write an essay critiquing it. Students would be rated on the performance and the critique. Or, they might

compare their own performance to a performance of the same music by professional musicians, answer written questions (perhaps multiple-choice) about the music, or evaluate the performance of peers. (For a more extensive description, see National Art Education Association, 1995.) Illinois began pursuing that approach as one of 15 states that worked with the Council of Chief State School Officers to develop exercises for the NAEP assessment. Later, several arts educators began developing exercises for use in Illinois.

Other Forms of Assessment

Other forms of assessment in the arts—for example, performance, portfolio, or essay assessment—may provide very important information about student learning in the arts. However, the Illinois state department of education has not begun to pursue them because, currently, they do not appear to be feasible for state assessment. Performance assessment may be the most valid approach for collecting information about student attainment of the state goal focusing on creation and performance in the arts. However, problems related to developing scoring rubrics and establishing interrater agreement are multiplied because that goal applies to each student in only one of the four arts areas, and the type of art may vary within each (for example, singing or playing an instrument; playwrighting or acting; choreographing or dancing; painting, sculpting, or filmmaking). Also, unless performances are scored immediately, they will have to be

videotaped.

Portfolios are probably most appropriate at the local level, where local criteria can be used to evaluate student work. Also, it will not be necessary to ship everything to a central location and then arrange storage. Portfolios can be especially valuable for showing student growth over time. (However, portfolios cannot be considered assessment tools unless they are evaluated systematically using a rubric developed in advance. Without that scoring, portfolios may be very effective instructional tools but do not assess student learning.)

Essay assessment might also provide important additional information about student learning. For example, it might more effectively assess students' ability to use sensory qualities or organizational principles of the arts at a higher cognitive level than multiple-choice tests. However, written essays provide information about student performance only if that performance requires students to write essays. They do not actually assess student performance in the arts.

The Limited Role of State Assessment in the Arts

The state assessment of the arts in Illinois will provide a general—though somewhat limited—snapshot of student learning in the arts. It will not provide information about students' ability to create or perform in the arts. It will provide some information about students' ability to apply arts concepts—for example, it will test their ability to recognize organizational prin-

ciples in music notation or in images of major visual artworks.

However, the state assessment will provide a great deal more information about student learning in the arts than is currently available. Furthermore, it will help focus the attention of educators and members of the public on what students are and are not learning in the arts. And, the state department of education is providing more resources (grants, workshops, technical assistance, prototypes, and other written resources) to help support local assessment than it could have otherwise.

The Role of Local Assessment

The most meaningful information about what students are learning in the arts can probably be collected at the local level. There, educators can decide what types of approaches will provide the most valid assessment of locally-important goals. Admittedly, developing and implementing such an assessment is not an easy task. It will require many resources, perhaps including training in assessment for teachers and others. The temptation to [latch onto] and use any assessment procedures that become available will probably be great. In the long run, however, schools may benefit most if they rationally and cautiously adopt and develop assessment procedures that are of high quality and clearly assess skills and knowledge that are important locally. This may mean starting small and building deliberately and gradually. It also means recognizing that assessment tools cannot be completely developed

in a limited period of time. The tools must be reviewed (preferably by people who were not involved in writing them) and tried out with students. Some may have to be revised and pilot tested again because they did not "work" the first time. Some may need to be revised again within a few years because, for example, local goals have changed or students are learning so much more that the assessments need to be written at a higher level.

Assessment developers must also remain vigilant about validity and reliability even though traditional methods of estimating validity and reliability may sometimes be inappropriate with new forms of assessment (Linn, 1995). However, it is critical that information from assessment can be used to make appropriate conclusions or decisions about arts learning and arts programs. Unless assessment data are of high quality, their credibility may be low. People will not take them seriously, and they will not help improve or gain support for arts education.

References and Other Assessment Resources

Consortium of National Arts Education Associations. (1994). National Standards for Arts Education. Reston, VA: Music Educators National Conference.

Herman, J. L., P. R. Aschbacher, and L. Winters. (1992). A Practical Guide to Alternative Assessment. Alexandria, VA: Association for Supervision and Curriculum Development.

Illinois State Board of Education. (1995). Assessment Handbook: A Guide for Developing Assessment Programs in Illinois Schools. Springfield, IL: Illinois State Board of Education.

Illinois State Board of Education. (1995). Effective Scoring Rubrics: A Guide to their Development and Use. Springfield, IL: Illinois State Board of Education.

Linn, R. L. (1995). Assessment-Based Reform: Challenges to Educational Measurement. William H. Angoff Memorial Lecture Series. Princeton, NJ: Educational Testing Service.

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

Linn, R. L., E. L. Baker, and S. B. Dunbar. (1991). Complex, performance-based assessment: Expectations and validity criteria. Educational Researcher, 20(8), 15-23.

Marzano, R. J., D. Pickering, and J. McTighe. (1993). Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model. Alexandria, VA: Association for Supervision and Curriculum Development.

National Art Education Association. (1995). NAEP Visual Arts Assessment and Exercise Specifications. Reston, VA: National Art Education Association.

Teaching and Assessing

Keith Swanwick

Living is assessing

Assessment guides all our actions and is usually unconscious, informal, intuitive. There is often no standard procedure to follow, no need to analyze, no need to write a report. Indeed, if we were to stop to quantify angles and speeds when crossing the street we would be at risk. Yet we certainly assess the situation. We see this informality at work in our daily judgments about music. For example, we might turn on the car radio and find that it emits a performance that we just do not want to hear. We switch off or we change the channel; filter the music out or in. At a simple level we will have made an assessment.

Assessment is an inevitable part of the fabric of teaching. Teaching involves responding to what students do and say. Responding suggests that we 'read' what is happening. We assess continually, interacting both with students' music making and with what they say about music. Formative assessment - the process of 'in-forming' - is an essential part of teaching and learning.

Teaching is assessing

The teacher points out and discusses the relationship between aspects of the music, querying anomalies, drawing attention to special strengths, and suggesting extra possibilities. He or she discusses what skills are needed for the task in hand, to what extent they have been successfully deployed, how they might be perfected, what further skills might more fully realize the music, and how these might best be acquired. The teacher tries to get the pupils to bring fully into play their own listening and self criticism, so that the process becomes an interaction between self-assessment and teacher assessment. This is assessment in the most educationally important sense of the word. (Loane, 1982).

A good music teacher is a good music critic. The first requirement of a music critic must be to acknowledge the complexity of musical experience. Such a rich activity cannot be reduced to a single dimension, say that of 'technique'. On the other hand, it does not make sense to identify several different dimensions and assess them giving a separate mark for each - say for technique, expressiveness and stylistic awareness - adding them up to get a single figure. When we conflate several observations we lose a lot of important information on the way. For instance, in competitive ice skating one performer might be given six out of ten for technique and nine for artistry, while another contender gets nine for technique and only six for artistry. The sum of each set of marks happens to be the same - 15 - but the actual performances will be quite different. The fudge of adding a category called 'overall' only makes things worse.

We must also resist falling back on the poor levels of meaning

embodied in numerical marks and we ought to beware of the false impression of exact quantification that numbers can give. Nor can we assume that more of something (say greater instrumental agility) necessarily means a better performance. For instance, we might be tempted to think that playing lyrical pieces by Grieg is of less musical value than throwing off a virtuoso delivery. Surely the person playing the 'easy' piece ought not to get such high marks as the other? This really is not satisfactory and may make too much of mere virtuosity, and may lure performers into water too deep technically for the good of their musical development.

There is a further problem. How are we to equate the technical levels of, for instance, a piano performance of a Chopin mazurka, a Scott Joplin rag and a Bach Fugue? There are different kinds of technical challenge here. While Bach requires clear and balanced part-playing with matching articulation for each appearance of the same material, both Chopin and Joplin need an accurately placed left hand and careful colouring of the chords in the inside parts that does not detract from the elaborate flow of the melody. It might be hard to say which piece is the most demanding to play; there are different difficulties. Yet in spite of obvious differences we can find some things in common between a musician playing one piece and someone else playing another; in the same way that it is possible to say of a well-known tennis player, an Olympic high-jumper and an international footballer that they are all fine athletes.

We do this by applying general criteria that define what it means to be athletic. In the same way, there are qualities that we recognize as musical wherever they appear. Can we identify these? If so then we are on the way towards declaring our criteria for musical assessment, towards putting our cards on the table. We have to articulate what counts as musical understanding.

The Dimensions of Musical Criticism

Salieri takes a first look at the score of Mozart's Wind Serenade in Bb:

"On the page it looked like nothing; the beginning simple, almost comic, just pulse - bassoons, basset-horns - like a rusty squeeze- box. And then, suddenly - high above it - an oboe, a single note hanging there, unwavering, until a clarinet took it over, sweetening it into a phrase of such delight. This was no performing monkey. This was a music I had never heard, filled with such longing, such unforgettable longing. It seemed to me that I was hearing the voice of God."
-From the film Amadeus, Peter Shaffer.

Salieri has an impression of particular sonorities from which he can instantly 'place', or label the music as simple or comic - the kind of thing he has heard so many times before. He can of course imagine the instrumental texture and its commonplace 'squeezebox' effect. But the high oboe note which evolves into a phrase on the clarinet transforms his impression. Sounds have become sequenced into expressive gestures. The score

is snatched away from Salieri before he can say more about the structural evolution of the music but he is left with a strong feeling of its value that his whole life is changed.

We might notice how on his first encounter with this music he tries to categorize it. Labelling is part of a filtering processes by which we select or reject on the basis of musical type. For instance, an eight year-old said of the introduction to Phil Collins singing That's Just the Way It Is, 'a pop star would sing this' (Hentschke, 1993). But like Salieri, she also notices the instrumentation - 'it was a drum'. A ten year-old goes further, telling us that the music had:

-wooden instruments at the be ginning and strings, had a drum beat,the wooden instruments at the beginning had a pattern and later it went faster, relaxing after the percussion instruments

She notices the expressive effect of changes in speed and is able to characterize this as 'relaxing'. She also is aware of the instrumentation - though it is keyboard rather than strings that we hear. Older children are likely to relate to expressive character even more strongly. Here a 13-year-old tells us that the beginning of this same recording:

-makes you feel the music rather than listen to it, because it makes you take notice of what was actually played - very nice and very 'flowy'. The second part was very bouncy and then comes the flowy part. I like this type of music best, because you can feel it, and (it) makes you

feel involved. It sounded like night, but when the drums come in it sounded like stones falling.

It seems taken for granted that drums are playing. The real point is that they change the level of expressiveness from 'flowy' to 'stones falling' - an engagement that lies at the centre of our understanding of musical form. For form is about relationships, not only between larger architectural blocks but more organically between gestures, patterns, phrases. These can be heard as repeated, transformed or contrasted, this can lead us on expectantly, sometimes surprising us - 'suddenly -an oboe - a single note hanging there'.

We can also see in these descriptions an emotional empathy. It 'makes you feel involved', or - more grandly - we are 'filled with such longing'. Reflecting on our own musical experiences, we know how our responses can grow into a strong sense of the value of music, a celebration of its human significance. All of these elements permeate musical assessment. These are the same components we saw in the fictitious comments of Salieri, dimensions of criticism that characterise the quality of musical engagement. We need to come clean over our criteria for these.

Ranking orders, numbers or league tables are not helpful forms of assessment. They tell us virtually nothing about the quality of the activity. This is why assessment procedures in education have moved towards 'criterion referencing', matching what students do with descriptions of different types and levels of activity. Criteria or critical descriptions' are supposed to help us structure our judgments and comments, but only if they are drawn up in appropriate ways. Criteria should have all the virtues listed below.

Criteria for criteria

They should be:

- clear;

- qualitatively different from each other;

- brief enough to be memorable but substantial enough to be meaningful;

- able to be hierarchically ordered;

- useful in a range of settings

- reflect the essential nature of the activity.

Relating a musical performance to a criterion statement is like matching our impression of someone to a photograph or painting. We see the picture 'whole'. In the same way we cannot break criterion statements into bits. Either there is a fairly good overall 'likeness' or there is not. When using criteria in musical assessment, a performance is placed in a particular qualitative category. Criterion assessment is dependent on the recognition of qualities, not on numerical quantity. The question is always the same: which statement most nearly fits this playing?

I can offer some criterion statements for musical performance. These have already been found helpful in adjudicating musical performance, though they arose initially from work in assessing children's compositions (Swanwick and Tillman, 1986). Since then they have been further developed and put to the test (Swanwick, 1994). In one check for reliability each was typed onto a separate card and the set of eight cards was shuffled into random order. Working in nine groups of three or four, 30 experienced music teachers ordered the cards into what seemed to be the most likely progression, from the statement reflecting the lowest level of performance to that describing the highest. There was a very high level of agreement within and between the nine groups of judges. It was possible to place the cards in rank order. These statements meet our conditions. They certainly seem to be meaningful and they are clear and qualitatively different from each other.

Eleven recorded performances for a public examination were played to seven teacher/judges. The performances included Domine Deus (Vivaldi), Beethoven's Fur Elise, a folk tune played on recorder, an ensemble playing the popular song, You are Always, and a specially composed piece played on electronic keyboards. Without discussion and quite independently, the seven judges reached a very high level of consensus about the relative merit of all the performances, different in kind though they all were. The least reliable judge was the eighth, the 'real' examiner who had finalized the marks but had not been working to criterion statements! Here are the criteria.

Criterion statements for musical performance

Level 1: The rendering is erratic and inconsistent. Forward movement is unsteady and variations of tone colour or loudness appear to have neither structural nor expressive signifiance.

Level 2: Control is shown by steady speeds and consistency in repeating patterns. Managing the instrument is the main priority and there is no evidence of expressive shaping or structural organisation.

Level 3: Expressiveness is evident in the choice of speed and loudness levels but the general impression is of an impulsive and unplanned performance lacking structural organization.

Level 4: The performance is tidy and conventionally expressive. Melodic and rhythmic patterns are repeated with matching articulation and the interpretation is fairly predictable.

Level 5: A secure and expressive performance contains some imaginative touches. Dynamics and phrasing are deliberately contrasted or varied to generate structural interest.

Level 6: There is a developed sense of style and an expressive manner drawn from identifiable musical traditions. Technical, expressive and structural control are consistently demonstrated.

Level 7: The performance demonstrates confident technical mastery and is stylistic and compelling. There is refinement of expressive and structural detail and a sense of personal commitment.

Level 8: Technical mastery totally serves musical communication. Form and expression are fused into a coherent and personal musical statement. New musical insights are imaginatively and systematically explored.

These have been very serviceable in all kinds of educational settings. They are especially helpful when we are not permitted the luxury of informal, intuitive assessment but have to report our assessment formally to students, parents, and the school administration. The analytical framework implicit in these particular statements has a research base and it appears to serve musical sensitivity as well as critical transparency. I hope that colleagues in the USA might try them out.

References

Loane, B. "The Absurdity of Rank Order Assessment." Music in the Secondary School Curriculum: Trends and Developments in Class Music Teaching. Ed. J. Paynter. New York: Cambridge U. Press, 1982: 242.

Hentschke, L. (1933) Musical Development: Testing a Model in the Audience-Listening Setting, unpublished PhD thesis, University of London, Institute of Education.

Swanwick, K. and Tillman, J. (1986) The Sequence of Musical Development: A Study of Children's Composition, British Journal of Music Education, Vol. 3, No. 3, Nov. 1986, 305-339.

Swanwick, K. Music, Mind and Education, London: Routledge, 1988.

Swanwick, Keith. Musical Knowledge: Intuition, Analysis and Music Education. London and New York: Routledge, 1994.

Planning for a Statewide Arts Assessment:The Western Australian Experience

Carol Myford

Opera Diva Dame Joan Sutherland, jazz musician James Morrison, composer Colin Brumby, aboriginal rock band Yothu Yindi, folk musician Kev Carmody, country and western artist John Williamson - Australia's musical scene is alive and diverse as its people. But what do children growing up in Australia today learn about music? What do they know? What are they able to do? Music educators are presently working in the Education Department of Western Australia to design a statewide assessment that will provide some answers to these important questions.

Before launching into the details of the assessment plan, a very brief geography/civics lesson might be in order to provide American readers with some context for understanding the project. Western Australia covers about one third of the continent- the largest of the six states and two territories that make up Australia. The state is over a quarter of the size of the US. (more than three and a half times as large as Texas). In this huge expanse live 1.7 million people; 1.2 million reside in and around Perth, the state capital, which is located in the southwest corner of Australia and is often referred to as the most isolated city in the world. In the rest of the state, inhabitants are more thinly spread out, residing mainly in coastal towns, with comparatively few living in remote areas of the state's vast arid interior. Over the last 30 years, Australia has become increasingly culturally diverse. Currently, there are 140 different ethnic groups represented on the continent, speaking 90 different languages.

About a third of the students in Western Australia attend private schools while the rest attend Government Schools (i.e. remote community schools, special education centers, the Distance Education Centre, district high schools, and both small and large country and metropolitan schools). There are 770 schools in the Western Australian state system serving approximately 250,000 students.

In 1009, the Education Department of Western Australia initiated the Monitoring Standards in Education (MSE) Project to assess that Western Australian students know and can do eight key learning areas: English, Mathematics, Science, Studies of Society and Environment, Health and Physical Education, Technology and Enterprise, The Arts, and Languages other than English. Since 1990, statewide assessments have been administered in English, Mathematics, Science, Studies of Society and Environment, and Health and Physical Education. Policy makers and education officials use the assessment results to monitor system-level performance in these key areas and to assist individual schools in monitoring their own students' performance.

The MSE Project assesses strati-fied random samples of students in Years 3, 7, and 10 (which correspond to Grades 3, 7, and 10 in the US) from the full range of Government schools. When drawing the samples, certain subgroups (e.g., Aboriginal students, students from non-speaking backgrounds) are purposely over sampled so the more precise estimates of their performance can be obtained. The Education Department carefully monitors the performance of these subgroups on the statewide assessments and uses the test results as a basis for initiating reforms at improving access to education for 'at-risk' students.

The Education Department reports student performance in a learning area in relation to a standards framework that contains student outcome statements describing the knowledge, skills, and understandings that the Western Australian Government school system believes are essential for all students to acquire. The outcome statements describe in progressive order eight levels of reporting that are meant to cover the compulsory years of schooling (Years K-10). These eight achievement level statements form a continuum along which students' performances can be mapped. (The achievement levels statements do not correlate with year levels).

In 1994, the Education Department published a preliminary working draft of a standards framework for The Arts. The framework contains outcome statements for five arts disciplines: dance, drama, media, music, and visual arts (art, craft and design). For each arts dis-

cipline, student outcome statements are presented in interrelated strands, or approaches to defining and understanding the arts. Initially, the framework identified three strands which, in the last year, have been expanded into six strands: (1) Exploring, Creating, and Developing ideas, (2) Using Skills, Techniques, and Processes, (3) Presenting, (4) Using Arts Languages, (5) Arts Criticism and Aesthetics, and (6) Past and Present Contexts.

This past summer, work began on the development of a statewide assessment in the arts that is slated to be administered in 1996—the first state arts assessment ever to be conducted in Australia. Jim Tognolini of The Educational Testing Centre at the University of New South Wales and David Andrich of Murdoch University in Western Australia jointly won the contract for the development of the assessment instruments and the analysis of student performance data. Beverly Pascoe, from the Education Department, serves as Project Officer for MSE while Annette Mercer, from Murdoch University, serves as Project Coordinator. Working groups of arts educators from the Government schools, MSE officers, arts consultants from Curriculum Branch of the state's Education Department and tertiary arts personnel have been meeting to design the blueprint for the arts assessment. In each of the five arts disciplines, there assessments will be built—one each for Years 3, 7, and 10. Each assessment will consist of two parts.

Part 1 of the assessment will be a paper-and-pencil analysis test that focuses primarily on as-

sessing student achievement related to outcome statements contained in the fifth and sixth strands of The Arts Framework (i.e., Arts Criticism and Aesthetics, and Past and Present Contexts). Students will be presented with several stimuli (e.g., for the music assessment, audiotaped excerpts of performances of choral and instrumental pieces). After viewing (or, in the case of recorded music, hearing) a stimulus, students will respond to a series of questions related to the outcome statements in the fifth and sixth strands. Some of these questions will be multiple-choice items, others will require students to write a short response (i.e., a word phrase, or perhaps a sentence or two), and a few may require more extended written responses.

Part 2 of the assessment will be a process portfolio built around a carefully circumscribed set of related tasks that are designed to take students several class periods to complete. Generalist classroom teachers or specialist teachers will administer Part 2 of the assessment to intact classes. Students in these classes will each prepare a mini-portfolio that has three components: (1) evidence of the planning process the student engages in when creating a product or performance. The portfolio activities will focus primarily on assessing student achievement related to outcome statements contained in the first four strands of The Arts Framework (i.e., Exploring, Creating and Developing Ideas; Using Skills, Techniques, and Processes; Presenting; and Using Arts Languages).

The challenges that the Western Australian assessment presents (i.e., designing, administering, scoring, scaling, and reporting on a large scale assessment in the arts) mirror many of the same challenges that we at the ETS are presently working through in conjunction with the National Assessment of Educational Progress (NAEP) arts assessment. In the month that I spent in Perth this fall consulting with MSE project staff and the working groups on their design blueprint, we found that we shared many of the same goals, for our assessments and a common research agenda. A number of the issues that we are confronting in the NAEP arts assessment are issues that the Western Australians will be struggling with as well. For example, listed below are a few of the research questions that our two assessment projects have raised for collaborative study:

-Do raters use scoring guides (referred to as "marking schedules" by the Australians) in certain formats more reliably than scoring guides in other formats? Are some formats easier to use (or to train raters to use) than others?

-When raters evaluate individual and group performances from videotapes, what difficulties do they encounter? How can these difficulties be overcome? Is it possible to generate both individual and group scores from these videotaped performances?

-Can we combine student's responses to multiple-choice items, rating of their written responses, and ratings of their performances or products in a single analysis? Can we scale

judged performances alongside more traditional test items? What are our options for combining and reporting results from the arts assessment?

-How shall we go about mapping the assessment results to achievement level descriptions?

What formulas for reporting assessment results will various audiences (i.e. governmental officials, principals, teachers, parents, the business community) find most useful? What information will each audience most need from this arts assessment?

The questions posted are formidable ones, indeed; but we at ETS are encouraged that assessment developers, psychometricians, and researchers on two continents will now be jointly wrestling with them. We look forward to nurturing and sustaining our intercontinental collaboration, sharing and learning from one another's experiences as we strive to make these assessments a reality.

If readers would like a request a copy of the working draft of the Western Australia arts framework or additional information about plans for the state arts assessment, they may contact Ms. Annette Mercer, Project Coordinator, MSE The Arts, Murdock University, School of Education, Murdock 6150, Western Australia or Ms. Beverly Pascoe, Project Officer, MSE The Arts, Evaluation Branch, Education Department, 151 Royal Street, East Perth, Western Australia 6004. Ms. Mercer can be reached by e-mail: mercer@murdoch.edu.au. Ms. Pascoe's e-mail address is: e448155@zexesupp2.eddept.WA.edu.au.