



## Special Research Interest Group in Measurement and Evaluation

### TABLE OF CONTENTS

Introduction and Invitation.....	1
Testing in MN.....	2
Assessments in NY.....	3
Lessons From IL.....	8
ME, MO, PA Reports.....	10

### INVITATION CAROL RICHARDSON Editor

Welcome back to the Measurement and Evaluation Newsletter! In case you were wondering, the newsletter has not appeared since 1998 so this newest edition brings us up to date on recent assessment developments in six states: Illinois, Maine, Minnesota, Missouri, New York, and Pennsylvania.

Measurement and evaluation are generating a lot of heat in the national news media, with issues ranging from grade-level assessments in core subject areas to basic skills tests for high school graduation to teacher competency tests for certification. The reports included in this issue reflect the continuing impact of politics on the assessment of student learning in music, the arts, and in music teacher education programs. It remains the task of thoughtful, forward-thinking leaders in music education research to initiate and shape the discussion at both the state and national levels, and help create the assessment vehicles that most accurately reflect both music learning and music teacher competencies.

It is in this spirit of initiating dialogue that we planned the SRIG session at Nashville. The session, titled "The Role of Assessment in Improving Undergraduate and Graduate Music Education Programs," brings

*(continued on page 11)*

### INTRODUCTION RICHARD COLWELL Editor, Emeritus

There is much of importance to this SRIG Newsletter in Measurement and Evaluation; perhaps of most importance is its resurrection by Dr. Carol Richardson and Dean Karen Wolff of the University of Michigan. These two individuals recognize the importance of the discipline of assessment in the health of music and music education, not only today but for the future. We welcome their long-term commitment to this venture.

Accurate feedback is critical to learning in most sub-disciplines of music: ensembles, music history and theory-the core of the education of professional musicians. I have yet to meet an applied music teacher or musicologist who is shy of low or high-stakes testing. One issue in assessment as well as purpose is the extent to which required music in the public schools is similar to professional music teaching and learning. As Paul Lehman has often queried: "Are we about education or entertainment?" Music is often cited as a model of authentic assessment, which is true of part of our program. Our guilt is more in lack of assess-

*(continued on page 11)*

## The Testing Scene in Minnesota

PAUL HAACK, Professor, Music Education

JEFFREY KIMPTON, Director and Professor, Music Education  
SCHOOL OF MUSIC, UNIVERSITY OF MINNESOTA

Teacher certification and student assessment in Minnesota continue in a state of transition. Both areas reflect the trends seen nationally, but also indicate differences of opinion about the role of testing within the state among school districts, professional organizations representing administrators and classroom teachers, and professional teacher organizations in the arts. In the mix are a governor and legislature who are also increasingly at odds over the rigor and purpose of testing in the state.

### PROSPECTIVE TEACHER TESTING

Minnesota continues to follow national patterns in teacher testing, although all teacher testing is limited to pre-service candidates. *Music* teacher testing is no different than that which is required in other curricular areas, with the exception of the Praxis II Content (Subject Area) tests. The State Board of Teaching is in the process of expanding the testing for licensure program, which has until very recently consisted of only Praxis I (reading, writing, and mathematics). The cost to prospective teachers in terms of registration and testing fees will be in the \$350 range. Candidates must pay the full amount for testing during the trial period as well as after. For the past year the University of Minnesota passing rate for Praxis I was 99%, a statistic that causes some to question the need for the test.

Those making application for a Minnesota teaching license prior to September 1, 2002 must:

- Take Praxis I and pass the three test portions.
- Take Praxis II: Content (no passing scores required prior to September 1, 2002)
- Take Praxis II: Pedagogy (principles of teaching and learning-- no passing scores required prior to September 1,

2002)

Those making application for a Minnesota teaching license on or after September 1, 2002 must:

- Take and pass the Praxis I PPST/CBT: reading, writing, mathematics tests
- Take and pass the Praxis II: content (subject area) test
- Take and pass the Praxis II: pedagogy (principles of teaching and learning) test

Minimum scores are as follows:

Praxis I: reading 173 or 320; writing 172 or 318; mathematics 169 or 314

Praxis II: music content (subject area) knowledge 140

Praxis II: pedagogy--grades K- 6 = 152; grades 5 - 9 = 144; grades 7 - 12 = 153

There are developing concerns over what these increasing series of benchmark tests ---and their costs---could mean for the pool of available teachers. The rural areas of the state are currently experiencing a significant shortage of teachers, especially in the sciences, special education, bilingual education, and now even in music. There is a growing resistance from entry-level teachers to teaching more than one hour away from the Minneapolis-St. Paul metropolitan area. Both are trends that bear further observation.

### K-12 MUSIC STUDENT TESTING

Student assessment for K-12 districts, part of the state's controversial portfolio and performance-based program *The Profile of Learning*, is under fire by Minnesota superintendents, teachers and legislators. Currently Minnesota's testing emphasis is on "core areas" defined as mathematics, reading, and written composition. Mathematics and reading are first tested at the 8<sup>th</sup> grade level with the

test taken annually until passed. Written composition is first tested at the 10<sup>th</sup> grade level with the test taken annually until passed. Districts may use the "state" tests, or tests of their own choosing, or tests of their own making, subject in the latter two instances to state review and approval. In any event, passes in all three areas must be attained for awarding of a high school diploma.

Minnesota also has "Comprehensive Assessments" in the same three areas at the 3<sup>rd</sup> and 5<sup>th</sup> grade levels. In the long range plan all learning areas, including the arts, are to be tested, and all of these tests are to be extended to the 8<sup>th</sup> and 11<sup>th</sup> grade levels. The Comprehensive Assessments are not to be used to keep students from graduating. Their primary purpose to verify what schools are reporting about their performance assessments with regard to the State's "Profile of Learning Standards." The Profile projects are by and large culminating efforts that are done while in high school, and those profile areas that a district *elects* to include are then required for graduation.

Assessment in the arts had been included with all other core standards since the *Profile* was adopted, but that part of the Minnesota testing program (along with foreign language, among other subjects) has been the most controversial in part because it requires so much one-on-one time and record keeping, is expensive, and requires arts teachers to assess *all* students in a school, not just those who elect to take music. Some teachers feel this portfolio process to be unmanageable, and with difficulties enough in meeting assessment goals in even the core areas, many districts have begun to seek a way out of the requirements in the arts and other areas. Some politicians believe the *Profile* is a novel assessment component that is popular with some constituents; others feel it is simply expensive and intrusive. This year the legislature decided to give districts new freedoms in selecting the comprehensive assessments, and at this time local school districts have some latitude in selecting those assessments that meet district priorities.

(continued on page 8)

# Developing Standards-Based Assessments in Music and Other Arts for New York State Students: The ASSETS Project

MARIA RUNFOLA, SUNY-BUFFALO

New York State has a longstanding tradition of state-developed locally administered academic assessments for its students. These assessments are referred to as The New York Board of Regents Examinations or, colloquially, "Regents Exams". The Regents Exams are part of a Statewide Testing System that includes other state assessments as well as local school district examinations. A student's scores on the Regents Exams determine whether he or she will graduate from high school early, on-time or at all. The scores also figure into a student's chances of being accepted into the higher education institutions of his/her choice.<sup>1</sup>

At present, New York State is developing Arts assessment models as part of its testing system, which schools may use as part of their local assessment. Once these models have been implemented by the state, schools will be encouraged but not required to use the Arts assessments. Unlike Math, Science and English components, a passing test score in the Arts component of the NYS Testing System will not be required for graduation. However, the integration of standards-based Arts assessments into the NYS Testing System reflects, in part, the state's commitment to Arts education and its recognition of the Arts as an important part of an education.

The New York State Board of Regents has adopted 28 Learning Standards that serve to guide educators in New York and four of these standards are in the Arts (see Figure 1.) In line with these standards, the state mandates that students receive instruction in the Arts, defined as dance, music, theater and visual arts, from grades Pre-K through 12. Since the mid 1980s, the Regents have required that in order to graduate from high school, students must complete one unit of credit in any one of the four arts disciplines. A

unit of credit is defined as 180 minutes of instruction per week or its equivalent and the mastery of course requirements in a State-developed or State-approved locally developed syllabus. The proposed integration of state-wide standardized evaluations both impacts and reflects national trends.

The present assessments being considered by the New York State Board of Regents have been designed by teachers to provide multiple ways of testing a student's learning in the Arts. All four arts assessments are similar in structure, consisting of three elements: an on-demand written test, an on-demand performance event, and a performance/portfolio of the student's work throughout the year. Each assessment contains items particular to that discipline. Throughout all stages of development, NYS teachers and students have been actively involved in shaping and refining the tests. In general, professional organizations see the State Initiative in Standards and Assessments as a window of opportunity to raise both the quality and quantity of Arts Education for students in New York State.

In 1998, an established Ad Hoc Committee comprised of school Arts teachers, higher education Arts specialists, leadership from within the professional organizations and representatives from the New York State Education Department met with Commissioner of Education, Richard Mills to discuss the possibility of State Arts Assessments. The Committee pointed out that New York State Arts educators had the background and professional interest to assess student achievement, and asserted that students deserved the opportunity to demonstrate the full scope of what they had learned in Arts classes. The Commissioner concurred and encouraged the Committee to begin development of standards-based

Arts assessments. The Ad Hoc Committee<sup>2</sup> members established a direction for the Project, provided on going advice, and reviewed progress. A discipline specific Taskforce was appointed for each of the Arts - Dance, Theatre, Music and Visual Arts. From year to year, we maintained the same overall representation of the original Taskforces, while bringing in new members, active in their respective professional organizations, to represent schools across the State.

With funding from the federal Goals 2000 program<sup>3</sup> we launched a 3-year research and development project, named ASSETS (Assessments, Standards, Staff Education & Technology Systems) in the Arts. Our initial efforts focused on developing high school level assessments related to the Commencement General Education Level (CGEL) performance indicators in the Arts Standards. Our findings (outlined below) reflect theory and practical research, as well as significant teacher/student input. The ASSETS Project offers a design for providing multiple, effective ways to measure student learning in the Arts for New York State, and a model for developing similar designs in other states.

## Standards-Based Assessments in the Arts: Theory and Practice

The designers of the proposed assessments faced several challenges, two of which were paramount to the success of the project. First was the great diversity of backgrounds in the Arts among students. As Boyle noted, "The scope of what might be achieved in music is great...and may include general musical knowledge, knowledge of notation, aural-visual skills, aural skills, performance skills, and composition" (1992). Similar statements could be made for the other Arts as well. Moreover, student experiences in the Arts vary considerably, because much of it is incidental and in-school contact with Arts education varies widely from place to place. The test had to accommodate students in general Arts courses (e.g. Music in Our Lives) as well as those students who were fulfilling the one credit graduation Arts re-

quirement by being in more advanced courses (e.g. Wind Ensemble, Studio in Arts, etc). Still further curricular variability is introduced because the State does not require specific curriculum to address the Standards; instead, it allows each school district to develop its own plan for meeting the Standards.

Next, we were faced with the challenge of selecting which assessment techniques to use. Because of the nature of the Arts, students' performances cannot be judged solely based upon written exams and there is great debate among educators as to the effectiveness of various techniques when evaluating students' artistic growth. Compounding these concerns, was the necessity to develop appropriate assessment instruments for each of the four Arts disciplines. Student skills in each discipline needed to be tested via measures on par with those of the other disciplines. Further, since the Taskforces felt strongly that assessments should come from exemplary student work, we needed to develop formal standards for determining academic excellence in the Arts.

The Taskforces drew upon the work of several influential assessment scholars. Nierman (1996) suggested including a variety of assessment techniques in performance assessment, such as projects, interview/oral presentations, demonstrations and portfolios. His advocacy of multiple criterion measures for evaluation of student behavior reflected the work of earlier scholars such as Metfessel et al. (1967), who advocated the use of self-reporting by students on perceived level of achievement, peer evaluations, standardized rating scales and checklists as means of evaluating students in "visual arts, crafts, shop activities, penmanship, creative writing, exhibits for competitive events, cooking, fashion design, and other activities." Metfessel et al. proposed a model of teacher-created rating scales "for observation of classroom behaviors, performance levels of speech, music, and art."

Observation of real-time performance has been the gold standard for evaluating musicians, dancers and actors as well as other disciplines in the world beyond the

school and therefore constitutes "authentic" assessment (Radocy, 1995; Wiggins, 1989). Critics of performance assessment charge that these techniques are not considered psychometrically sound primarily because of a lack of evidence for the reliability and validity of these techniques. Inter-rater reliability, the largest area of concern, has improved and continues to do so as the data pool becomes larger, more studies are done, and methods of training raters are refined<sup>4</sup>. The constant re-combinations of raters strengthens reliability through constant re-examination of what should count as benchmarks in student works of art, performance, criticism, aesthetic decisions, and reflective writings. As the conversation broadens, a common agreement on terminology and levels of achievement emerges, bringing with it more valid and reliable results.

Accordingly, the Taskforces concluded that combined with other forms of measurement, performance observation in the Arts takes advantage of nearly 70 years of experience and research in its educational application and that indeed it was appropriate for these techniques to be included in the State assessments. However, we also realized that we needed a balance of the various assessment tools available and decided to include paper/pencil type techniques as appropriate to the various content areas (Colwell, 1970).

From the extant research, we generalized three logical possibilities for estimating reliability and validity of performance-based Arts assessments:

- 1) Teachers can award ratings to the same students who were tested on the same criteria under similar conditions on two occasions;

- 2) The same teacher, listening twice to recordings of students' performances or viewing twice Arts products and/or videotaped Arts performances, awards two sets of ratings to the same students who were tested on the same criteria on one occasion; and

- 3) Different teachers award ratings to the same students who were tested on the same criteria on one occasion.

Only the last of these was considered a

likely large-scale assessment scenario and thus became the procedure used in Scoring Sessions following each field test. During rubric development, The Taskforces felt that the assessments would work best if whenever possible, we would use measures which already existed and were familiar to the arts educators. For example, we took the contest rating sheets of the New York State School Music Association (NYSSMA) and with their permission adapted the sheets so that they could be used to assess the music performance events included in the music assessment.

## THE "ASSETS" PROJECT

The ASSETS Taskforces initially met to formulate assumptions for the project and to develop procedures for meeting their goals. The Taskforces established a set of ideal characteristics for a State High School CGEL Arts Assessment and the prototype tests were based upon the following assumptions: The assessments would be:

- State developed but locally administered.
- Based on the NYS Learning Standards for the Arts.
- Developed for each of the four Arts disciplines and would be reportable in the same format.
- Administered to all students.
- Reported to the State for public dissemination (including the state's existing reports on each district, known as the "School Report Card").
- Used as part of the requirements for passing the required arts course based on local district policy.

Item writing procedures were developed through the following processes:

1. Sample exams used by teachers across the state (urban, suburban, and rural districts) were collected;
2. Former State Education Department exams in the Arts, as well as other subject areas, were reviewed;
3. Texts appropriate for secondary level arts instruction were examined in order to develop a list of common repertoire, artists, terms, etc.;
4. Published standardized tests were reviewed for format, scoring, item types,

content, etc.

5. Examples of student performance collected throughout the State were examined and analyzed.

Next, discipline specific consultants<sup>5</sup> of national reputation in arts assessment, were brought in to work with the Taskforces and an overall Assessment Consultant<sup>6</sup> was hired to coordinate the various aspects of the project including production of all materials necessary for the field tests.

### ASSETS High School Arts Assessment Prototype

We then developed and tested item types using the following draft structure:

Part	Dance	Music	Theatre	Visual Arts	Number of Items	Administration Time in Minutes
<b>I. On Demand - written</b>	<b>40</b>	<b>45</b>	<b>50</b>	<b>45</b>		
A. Selected Responses(s)	20	30	20	20	20-30	1 each
B. Short constructed Response(s)	10	5	10	10	2	3-5
C. Extended Constructed Response	10	10	20	15	1-2	10-30
<b>II. On Demand - Performance Event (artistic)</b>	<b>25</b>	<b>10</b>	<b>15</b>	<b>20</b>	<b>1-2</b>	<b>10-50</b>
(On Demand Total: Two Hours)						
<b>III. Performance/Portfolio</b>	<b>35</b>	<b>45</b>	<b>35</b>	<b>35</b>		
A. Products (performance/works)	30	40	30	30	2-3	over time
B. Reflection (summary)	5	5	5	5	1	
<b>Total*</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	

\*Numbers represent proposed percentages of a score in each of the component areas and anticipated administration time per item type. Total maximum score is 100.

The music assessment contains 30 multiple-choice questions worth 30% of the student's score, two short constructed response items for 5%, and an extended constructed response item worth 10%. A sample multiple-choice item would show students a piece of music and ask them several questions about it. Other items have students listen to a musical selection and then select an appropriate response or write about the elements of music that are present.

The on-demand performance is unrehearsed. Materials are given to students to be reviewed and performed during a specific time frame, thus it is a timed event. For music, this performance consists of singing or playing a short musical composition after a brief study period. The on-demand performance event in music counts for 10% of the student's score.

The performance/portfolio requires one "centerpiece" or major task (a solo performance, an improvised solo performance, or a composition with performance); a repertoire list, a research project and a portfolio reflection.

After the item tryouts in the first year, we found the following:

1. It is possible to develop assessment items in all four Arts, since students responded to the various item types included.

2. Where instruction took place, students were able to perform the tasks addressed by that instruction. (e.g. If composition was a part of the curriculum, students responded appropriately to those items; where composition was not part

of the curriculum, student response was spurious).

3. Some tasks worked well, and some did not work at all.

4. Items ranged from fairly easy to very difficult, with most in the moderate to very difficult range. These item analyses assisted in identifying content difficulty levels across the state.

5. Students did well in districts where standards-based instruction occurred.

These results provided the basis for revisions and additions for the second year of development and tryout. Several items did not function appropriately because of administration logistics or other reasons. For example the music performance event, modeled after traditional "sight reading" procedures, was a dismal failure in part because general music students are not use to sight singing and/or sing-

ing alone, let alone recording their sight reading performance. For the second tryout, we compared two different models of the On Demand Performance Event: one with an easier difficulty level but the traditional "sight reading" procedure (the item was called "Sight Reading"); the other with the same difficulty level as the first year but allowing students five minutes to "practice" portions of the example before recording their performance (the item was called "Music Reading").

Student and teacher comments on the test items also assisted us in developing the second year revisions. For example, students in every discipline expressed preference for items that were based on prompts. This may be in part because they were used to Document Based Questions (DBQ's) in other subjects, or because of the "more authentic" nature of such items. In addition, we noted that items based on prompts generally measured higher order learnings rather than "knowledge" level and thus a decision was made to include more items with aural/visual prompts in the next round of tryouts. The wide-scale incorporation of high quality aural and visual prompts necessitates an assessment that is more expensive to produce, but which is also more effective and sustainable, and truly reflects the Arts disciplines. However, the fact that we could test much more material in an authentic manner with paper/pencil tests than originally anticipated made the on-demand written part of the test more "usable" i.e. scoring objectivity and accuracy could be increased while the amount of time needed for scoring could be reduced, especially if we designed the test to be machine scoreable.

In Year Two, because of notable revisions to the content of the assessments and improved application of the Learning Standards in participating districts, student scores on individual item types improved dramatically. Year Two results provided the basis for a complete assessment, and sample items were released to school administrators and arts educators throughout the State for the first time.

In its third year, the Taskforces established architecture for two operational

forms of the assessment for each of the Arts and delivered these architectures to the State Education Department. Two operational forms for each Art were finalized and also delivered to SED, and another tryout of the revised performance/portfolio section was released. The Taskforces developed Scoring Guides (including rubrics, directions for use and representative student responses to each item) for each discipline. The Scoring Guides were tried out during central scoring sessions in Albany. Based on the results of this third tryout, the Committee coordinated professional development seminars to be used in school districts statewide.

As part of these efforts, the State Education Department sponsored a broadcast on PBS "Tools for Schools" program and then released the second test sampler<sup>7</sup> including items, rubrics and samples of student responses to the sample test items.

The Arts Assessment Policy Committee then met with State Education Department officials, who reiterated their support for Arts assessment. SED field-testing was postponed for one year due to funding and staffing shortages but a more complete tryout of the curriculum embedded items for Part 3, the performance/portfolio, was accomplished. The State Education Department, with the Policy Committee's input, affirmed that participation in the first official assessment administration would be voluntary by district. Once the 2000-2001 performance/portfolio pilot exams were scored, ASSETS Taskforces concluded their work and the State took control of developing standards-based assessments for its students in the Arts.

In 2001-2002, full field tests are occurring at State Education Department-selected sites (representing the state's demographic spread). Thus, in the 2002-03 school year, districts will have the opportunity to participate in the actual assessments for the first time.

## CONCLUSIONS

Detailed analysis of the performance indicators in the New York State Arts Stan-

dards reveals that paper-pencil tests are appropriate measures for approximately 40% of the content implied in the Standards. All the discipline-specific Arts Assessment Taskforces recognize the value of written tests, and have proposed 40-50% of the Total State Assessment Score be derived from an on demand component that is primarily a paper-pencil strategy. The remainder of the Total Score would be derived from performance assessment and performance/portfolios developed by students over time. We arrived at this decision in part from surveying Arts teachers regarding the relative importance of the four Learning Standards as represented in their curricula. From the results, we concluded that Standard 1 was considered most important by Arts teachers, then Standard 3, and to a much lesser degree, Standards 2 and 4.

Arts educators can take advantage of several standard ways to increase reliability and validity of assessment scales. First, it is important that all dimensions and criterion statements for rating instruments (including rubrics and checklists), be directly related to the relevant performance indicators, in the Learning Standards as well as the learning processes that reflect common practice. Next, scoring criteria must be specifically defined for the students in the written or verbal directions for each item. Further, when constructing rating instruments, we should use both types of rubric scales available to us: holistic and analytic. This affords flexibility in design of the rubrics to accommodate the various skills being measured.

Finally, the new assessments must be field tested with a subject sample representative of those students who will eventually be assessed with these instruments. The rubrics must also be field tested to see if the statements developed to represent the various categories of Novice, Competent, Proficient and Distinguished are interpreted similarly by different teachers. This is most easily accomplished through centralized scoring sessions that include a staff development component as well as the scoring of all student responses by representative teams

of scorers.

With regards to Arts Education in general, Nierman and Metfessel's recommendations are congruent with the types of extended tasks proposed by the Taskforces for the CGEL statewide assessments in the Arts. The proposed assessment is comprised of a combination of item types from the traditional written multiple choice and essay questions as well as tasks which reflect authentic assessment practice in the Arts. Students complete required curriculum-embedded performance tasks and additional items to include in a performance/portfolio and through written reflection, each student justifies why those products exhibit his or her attainment of the Learning Standards. Current research points to performance tasks and portfolio assessment as viable and equitable forms of assessment for the Arts. Furthermore, ongoing research by a number of Arts Education scholars demonstrates repeatedly that it is possible to construct such assessment instruments in ways that yield high reliability coefficients, a necessary condition for establishing content validity.

As with any introduction of a new form of large-scale assessment, early efforts to establish reliability and validity for performance and portfolio assessment should be of primary concern. Some educators are suspicious of these activities because of their inherent subjectivity. These educators advocate confining assessment of student progress in the Arts to paper-pencil achievement tests. The trouble with this perspective is that it results in limiting assessment in the Arts to a very narrow range of content actually covered in instruction and cannot adequately assess content required by the Standards. Isaac and Michael (1995) caution that "when an outcome measure is conveniently available or has apparent *face validity* [emphasis mine], that measure becomes so well established that to depart from it challenges the credibility of the evaluation." Although it is true that traditional paper-pencil tests have the confidence of educators because of their *face validity* and relative ease of reliability estimation, we should not allow our choice of assessment technique to impede

the content validity of Arts achievement measures. Unfortunately, we do so if we maintain the position that only a paper-pencil test can reliably and validly measure student achievement in every discipline, including the Arts.

The New York State Arts Assessments are unique in a number of ways. First, separate assessments of equal value have been assigned for each Arts discipline. Also, students are given considerable choice in several sections of the assessment, which affords them opportunity to demonstrate achievement based on their own strengths and interests. In each assessment, students develop independent solutions to common tasks, and also select additional items they want to put in their portfolios. The assessments were designed so that a student cannot fail the entire exam by failing one part of the exam. Finally, the State's emphasis on performance and portfolio as methods to assess learning in the Arts allows for focus on the artistic process as well as product in each art form.

The assessments do not target particular course offerings, but focus instead on the Standards as a unifying goal. The assessments will be given as students complete Arts courses fulfilling the required unit of credit for graduation, achieving parity for the Arts with the other disciplines and offering the potential for the Arts to be included on the State Report Card.

Student success on the State Arts Assessments will require a coordinated curricular approach with students' intended learning outcomes mapped to the Learning Standards. Thus, the responsibility for success does not rest in the hands of the teacher who is in charge of the required course for graduation. Rather, it is the responsibility of all Arts teachers at every level to align their curricular expectations and content in order to prepare students for Standards-based assessment.

## References

- Azzara, C. (1992). Audiation-based improvisation techniques and techniques and elementary instrumental students' music achievement. *Journal of Research in Music Education*, 41 (4), 328-342.
- Boyle, J.D. (1992). Evaluation of music ability. In R. Colwell (Ed.) *Handbook of Research on Music Teaching and Learning*. New York: Schirmer, 247-265.
- Colwell, R. (1987). Test item difficulty and perception. *Bulletin of the Council for Research in Music Education*, 91, 19-22.
- Colwell, R. (1970). *The Evaluation of Music Teaching and Learning*. Englewood Cliffs, NJ: Prentice-Hall.
- Colwell, R. (1969-1970). *Music Achievement Tests*. Chicago: Follett.
- Gordon, E.E. (1971, 1991). *Iowa Tests of Music Literacy*. Chicago: GIA.
- Gordon, E.E. (1986). Final results of a two-year longitudinal predictive validity study of the *Instrument Timbre Preference Test* and the *Musical Aptitude Profile*. *Bulletin of the Council for Research in Music Education*, 89, 8-17.
- Gordon, E.E. (1967). *A Three-year longitudinal predictive validity study of the Musical Aptitude Profile*. Iowa City: University of Iowa.
- Isaac, S., & Michael, W.B. (1995). *Handbook in Research and Evaluation*. San Diego: Edits.
- Jones, H. (1986). An application of the facet-factorial approach to scale construction in the development of a rating scale for high school vocal solo performance. Doctoral dissertation, University of Oklahoma, 1986. Unpublished doctoral.
- Metfessel, N.S. et al, (1967). A paradigm involving multiple criterion measures for the evaluation of the effectiveness of school programs. *Educational and Psychological Measurement*, 27(4), 931-943.
- Nierman, G.E. (1996). Assessment-Applying the yardstick to musical growth. *The Nebraska Music Educator*, 55(1), 28-29.
- Radocy, R. E. (1995). Planning assessment in music education: It's not risk free. *The Quarterly Journal of Music Teaching and Learning*, 6 (4), 19-25.
- Runfola-Hale, M. (1977). An experimental analysis of the comparative effectiveness of harmonic and melodic accompaniment in singing as it relates to the development of a sense of tonality. *Bulletin of the Council for Research in Music Education*, 53, 23-30.
- Rutkowski, J. (1998). Validation of the 'Singer Accuracy Measure'. Presentation at the national meeting of the Music Educators National Conference, Phoenix AZ.
- Rutkowski, J. (1997). The nature of children's singing voices: Characteristics and assessment. In *Proceedings from "Sharing the Voices: International Symposium- 'The Phenomenon of Singing'"*. Memorial University of Newfoundland, Canada, 201-209.
- Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly*, 1 (1&2), 81-95.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46 (7), 141-147.

FIG. 1

## NEW YORK STATE LEARNING STANDARDS FOR THE ARTS\*

### Standard #1 - Creating, Performing, and Participating in the Arts

Students will actively engage in the processes that constitute creation and performance in the Arts (dance, music, theatre and visual Arts) and participate in various roles in the Arts.

### Standard #2 - Knowing and Using Arts Materials and Resources

Students will be knowledgeable about and make use of the materials and resources available for participation in the Arts in Various roles.

### Standard #3 - Responding to and Analyzing Works of Art

Students will respond critically to a variety of works in th Arts, connecting the individual work to other works and to other aspects of human endeavor and thought.

### Standard #4 - Understanding the Cultural Dimensions and Contributions of the Arts.

Students will develop an understanding of the personal and cultural forces that shape artistic communication and how the Arts in turn shape the diverse cultures of past and present society.

\*SOURCE: *New York State Department of Education.*

### Footnotes

<sup>1</sup> The Regents Exams were modified in the mid 1990's as part of a movement for national reform. The Presidential Task Force Report on public education, "A Nation at Risk" spurred simultaneous national state efforts directed toward establishing performance-based assessments (standardized assessments) for students. The "standards movement" was a response to the perception that the schools were not preparing many of their students to be successful and/or productive members of society upon graduation. NYS Commissioner of Education Regulations Part 100 require that the NYS Learning Standards be used to frame curriculum and that students be assessed on the knowledge and skills specified by the Standards.

<sup>2</sup> This Ad Hoc Committee became known as the Arts Assessment Policy Committee.

<sup>3</sup> Recently, the Goals 200 program was discontinued by the federal government.

<sup>4</sup> Rutkowski (1998, 1997, 1990) Azzara (1993), Jones (1986), Hale[Runfola] (1977), Gordon (1967), are among those who have achieved high estimates of inter judge reliability in studies they have conducted where the criterion variable was the result of using rating instruments to evaluate real time performance. Scoring procedures in these studies engaged teacher/raters in the assessment process utilizing multiple judges, panels, or readers. This is consistent with procedures characteristically used in Arts assessment.

<sup>5</sup> Dance, Susan Koff; Music, Richard Colwell; Theater, Karen Kay Husted; Visual Arts, Frank Phillips.

<sup>6</sup> Edward Roeber.

<sup>7</sup> The Sampler is available on line at <http://www.emsc.nysed.gov/ciai/arts.html>.

---

*The author acknowledges the following individuals for providing information to write this article: Edward Marschlok, Associate for Music & Theater, State Education Department; William Mercer, Arts Assessment Policy Committee; J. Terry Gates, Special Consultant for Goals 2000 Grant; Katy-Neyerlin-Colletti, ASSETS Coordinator; Jen Childress, Visual Arts Taskforce Chair.*

---

(continued from page 2)

This move clearly puts student assessment in the arts in a weaker configuration. In those districts that have chosen to include the arts in their Profile activities, most students choose to do their project in music. Some music educators believe that, were the profile to regain its original mandate, the number of students they serve would triple to accommodate those needing assessment in music. While there was euphoria among arts educators when the arts were originally included as a mandatory part of the *Profile*, there are now mixed emotions as to whether they are a boon or a bane to the programs of instruction in music in Minnesota schools.

## Lessons from Illinois

JUDY BUNDRA,  
SCHOOL OF MUSIC  
DePAUL UNIVERSITY

"More standardized testing is implemented in the United States than in any other country." ---Brandt (1985, p. 3)

We are a nation enamoured with numbers. In a world where the value of a politician is measured by opinion polls and the fate of a television series is determined by Nielsen ratings, numbers play a significant role in our everyday lives. The educational world is also filled with numbers--test scores have become a dominant measure of educational success or failure. Numbers are used to make quick and easy comparisons, pupil-by-pupil, district-by-district, state-by-state, and even country-by-country, and test scores are shaping the teaching-learning experience. As McColskey and McMunn (2000) stated, "High stakes testing--tests with important consequences for educators and students--have become the accountability tool of choice in many states as policy makers struggle to find ways to increase student achievement" (p. 115).

Where are the numbers for music education in the midst of this high-stakes testing? What tests are being administered in music education, and what have we learned from the tests? When the Illinois education reform movement began in 1985, the fine arts were named as one of the six fundamental learning areas of a K-12 education. But in 2002, the reality is far different than the reform rhetoric, and the arts remain at the fringes of Illinois assessment efforts. In this article, the role of music assessment in Illinois will be reviewed, and the status of of Illinois assessment will be analyzed.

### Illinois Education Reform Efforts in the 1980's

On February 10, 1980, the Illinois State Board of Education Goal Statement for the Arts was adopted as follows: "The arts should be viewed as an integral part



of the curriculum and every school system should assure that all students have access to exploration and study of the arts throughout their formal education” (Illinois State Board of Education, 1986, p. 1). Through extensive advocacy efforts by the arts education community, the fine arts were included as part of the education reform legislation passed by the Illinois State Legislature in 1985. State-wide assessment followed, known as Illinois Goals Assessment Program (IGAP) tests, which were administered to a sampling of children. In schools selected through a stratified random sample, students in grades 3, 6, 8, and 10, took a battery of tests in music, visual arts, dance and drama, comprised of 52 multiple choice items. Photographs of dancers and visual art works were included with the test. Although IGAP results were reported to the public in the areas of reading, math, science, and social studies, test scores in the fine arts and health/physical education were reported back only to the school districts. In “report cards” sent to the public, the state did not require school districts to disseminate fine arts and health/physical education scores. Because parents, students, and even teachers did not receive test results in fine arts and health/physical education, the impact of those IGAP tests was minimal.

In addition to the IGAP tests, school districts were also expected to write learning objectives and develop local assessment tools to be used at the district level. With little or no background in test development, music educators were expected to design their own tests. Student data had to be collected and stored, and teachers had to determine how many students exceeded, met or failed to meet district expectations. To oversee this entire process, the Illinois State Board of Education (ISBE) put a system in place known as “Quality Review,” which involved state visitations of local districts.

Reaction by the music education community to assessment mandates by the ISBE was largely negative. Miscommunication and policy changes by the ISBE led to misunderstanding and confusion, and music educators were reluctant to become involved in the testing move-

ment. Efforts to develop state and local assessments were hampered by limited time, funding, and resources. The ISBE faced rising demands from the public, yet sufficient dollars and personnel were not available to meet the expectations. Only one staff member was assigned the task of coordinating state-wide assessment in all four fine arts areas: music, dance, drama, and visual arts. The ISBE and arts organizations such as the Illinois Alliance for Arts Education (IAAE) and the Illinois Music Educators Association (IMEA) offered assistance to teachers through workshops and written materials, but teachers were largely unprepared to handle the state and local assessment tasks. Ill-equipped, untrained, and often times unwilling, many music educators quickly became frustrated by the ISBE mandates to assess student achievement.

#### **A Move to Learning Standards in the 1990's and 2000's**

The frequent testing in all six learning areas became onerous to both teachers and students. Faced with a complicated schedule of state and local tests, teachers found themselves spending less time on teaching and more time on testing. A Superintendent's Committee on Testing recommended major changes in the assessment plan, and the General Assembly passed legislation that modified the IGAP testing program. In 1993, a new schedule was implemented, with reading, math, and writing tests at grades 3, 6, 8, and 10, and science and social science at grades 4, 7, and 11. In conjunction with the social studies test, all 4<sup>th</sup> and 7<sup>th</sup> grade students were given questions about the fine arts, but not all students received the same set of questions. Fourth and seventh-grade students answered 10 questions each in two of four fine arts areas (dance, drama, music and visual arts), from a pool of eight sets of questions. This form of sampling was to yield data on all four fine arts areas, which could be compared only at the school or district levels. At the high school level, assessment in the fine arts became voluntary.

In 1997, the Illinois States Goals for Learning were revised and the Illinois

Learning Standards were adopted. Unlike the goals, which were broadly stated, learning standards were more specific in detail and in sequence. In order to reflect the new standards, another test known as the Illinois Standards Achievement Test (ISAT) was developed. In 1999, the first ISAT tests took place in reading, mathematics, and writing, and by 2000, science and social science were added, with music as part of the social studies test. At the high school level, the IGAP tests were replaced by the Prairie State Achievement Examination (PSAE), which was first administered in spring 2001 to 11<sup>th</sup> grade students. The fine arts are not part of the PSAE, although optional fine arts questions are available. In the summer of 2001, cut scores on the fine arts test questions were established, and a team of specialists and teachers determined what met, exceeded, or failed to meet state standards.

#### **What Are the Lessons from Illinois?**

Throughout the last 20 years of educational reform, with its accompanying mandates to test school children, music educators faced a number of difficult challenges. What can be learned from the state of Illinois?

**1) Valid assessment requires well-designed curriculum, and music educators must address the difficult questions of what should be taught, when, and how.**

Too often, assessment is not linked with the curriculum, and the curriculum is not linked with the standards. The proverbial tail is wagging the dog--rather than allowing assessment to grow out of instruction, instruction is being designed around assessment. Or worse yet, the dog has no tail at all, for in some schools, testing is done in the absence of any curriculum whatsoever. In high-stakes assessment, teachers find themselves teaching to the test, developing an educational agenda by default. If meaningful assessment is to take place, a balanced, comprehensive, and sequential curriculum is needed.

**2) Music teachers need to become in-**

**formed participants in the assessment process, with adequate time, training, and resources made available at the local levels.**

How can music teachers be expected to assess student achievement without sufficient planning time, training, and materials? Without time, fundamental issues surrounding music curriculum and assessment cannot be addressed. Without training, music teachers are unable to design an effective test or interpret the scores. Without resources, teachers cannot effectively measure or improve student achievement. These ingredients are foundational to meaningful assessment.

**3) The Illinois State Board of Education needs sufficient funding and personnel to design and administer effective tests, analyze data, and disseminate the results to the educational community and the public.**

The lack of a fine arts assessment coordinator in the ISBE offices is further testament to the marginalized status of Illinois fine arts assessment. If the arts are to be assessed state-wide, why is there no arts education specialist assigned to oversee the process? If the Illinois State Legislature declared the fine arts as a basic learning area, why is the test administered through sampling while other content areas test each student? Why are fine arts results unavailable to the parents, students, and teachers? How can the music tests impact music teaching and learning if the testing is done in a vacuum? Without adequate funding and personnel, the ISBE cannot support a valid assessment program in the fine arts.

**4) Test formats should be expanded, and alternative assessment, including portfolio or other performance-based assessment, should be considered a viable component of state-wide assessment efforts.**

The term "portfolio" is borrowed from the visual arts, yet it is ironic that the portfolio assessment project in Illinois does not yet include the fine arts. Although music teachers are encouraged to use performance-based assessment at the

local levels, efforts to assess music students using alternate forms of assessment need to be encouraged further at the state level.

It should be noted that the ISBE did, over a five-year period, attempt to design paper and pencil assessment in the fine arts using videotaped examples in dance, drama, music, and visual art as prompts. Although the test still was administered in a traditional multiple-choice format, the test itself contained artistic performances or representations in all four art forms. The videotape test was designed by a group of arts educators and ISBE staff members and then piloted at various schools throughout the state. Despite positive response from both teachers and students, the test was never distributed, and the project was placed on a shelf, largely due to cost and copyright laws. How effectively can we measure musical achievement in absence of sound? Not only should the assessment efforts include alternative forms of measuring student achievement, but the format of the traditional tests also needs to be expanded.

**5) A major shift in the public attitudes towards the importance of the arts is needed for meaningful assessment in music.**

Elliot Eisner (2000) noted the following: "Test scores drive curriculum because what is tested is what is taught. And since the arts are not tested, they can be neglected with greater immunity than those fields that are" (p. 5). This statement is certainly true in Illinois. When all Illinois students began taking the IGAP tests in music, administrators paid attention to the arts and arts instruction. As fine arts testing in Illinois became increasingly marginalized, the fine arts became more expendable. As LeBlanc (1986) observed, "Any branch of the school curriculum that fails to embrace testing will incur political risk through its failure to do so" (p. 37).

**What is the Next Step?**

One Illinois music educator called the public's preoccupation with testing "just

a passing phase." To the contrary, state-wide assessment continues to be a persistent, prevalent, and political issue. The role of assessment in Illinois demands serious attention from the music education community--we need a proactive, informed leadership that will help teachers deal with assessment in meaningful ways. As one administrator said, "If you don't help us, we'll mess it up for you." Will the music education leadership focus upon assessment? Will they advocate for better forms of assessment, more training, resources, personnel, and wider dissemination of results? Will the music education community work together toward change? Can our profession support the efforts of music teachers to use assessment to improve teaching and learning? Our students deserve nothing less.

#### *Bibliography*

- Eisner, E.W. (2000). Arts education policy? *Arts Education Policy Review*, 101, 4-8.
- Illinois State Board of Education. (December 1986). *State Goals for Learning and Sample Learning Objectives (4M 7-478B-26 No. 238)*. Springfield, IL: State of Illinois.
- LeBlanc, A. (1986). We must know what we really accomplish. *Design for Arts in Education*, 88, 37-39.
- McClosky, W. and McMunn, N. (2000). Strategies for dealing with high-stakes state tests. *Phi Delta Kappan*, 82, 115-120.

---

#### **OTHER REPORTS:**

##### *Maine*

As a result of recent legislation (LD 1760) Maine Schools will be required by AY '05/06 to implement the new Maine Learning Results. Music assessment will most likely become incorporated into the

Maine Education Assessment (MEA) battery, administered at the end of the 8<sup>th</sup> and 11<sup>th</sup> grades. They are being formulated by the Maine Dept. of Educational and Cultural Services, in conjunction with Maine Music Educators. They will be locally administered and will be performance assessments.

*Louie Hall  
University of Maine*

### **Missouri**

The Missouri Department of Elementary and Secondary Education and CTB/McGraw-Hill have developed a fine arts assessment for all Missouri students in grade five. The fine arts assessment is a component of the Missouri Assessment Program.

The Outstanding Schools Act, adopted by the Missouri legislators in 1993, called for increased accountability in improving student academic performance for all of Missouri's public school districts. In addition to creating the Show-Me Standards, a set of 73 rigorous standards intended to define what students should know and be able to do by the time they graduate from Missouri's public schools, the OSA addressed creating curriculum frameworks, a statewide assessment program, professional development for educators and professional standards for new educators.

The Missouri Assessment Program (MAP) is the statewide assessment system intended to provide an indication of how well students are meeting the Show-Me Standards and how well they compare academically with other students across the nation. Teachers, school administrators and community leaders were involved in all aspects of the development process, including the Show-Me Standards and the Framework for Curriculum Development in Fine Arts. Additionally, educators throughout the state have participated in item writing, content and bias reviews, and scoring the assessment. In August 2001 educators and community members were directly involved in setting the achievement levels and writing benchmark descriptors. The fine arts component of the MAP was piloted in 1999, field tested in 2000 and a volunt-

ary assessment was administered in 2001. The fine arts component is a video assessment. Students watch and/or listen to the stimulus, then answer questions in a student booklet. There are 40 selected response items in Session I, which is approximately 60-minutes in length. Session II, a 20-minute segment, is comprised of 5 constructed response items. Students are assessed in four areas: music, visual art, theatre, and dance. The mandatory assessment, scheduled for 2002, has been delayed until 2003, due to financial concerns in the state.

Music and art educators throughout the state have participated in professional development workshops designed to provide assistance in instructional strategies to enhance student performance on the Fine Arts MAP. The response from those teachers has been generally positive. Fine arts specialists understand the importance of being included in the statewide assessment program.

Our challenge in Missouri at this time is to ensure that the Fine Arts MAP will be reinstated in 2003. Arts organizations, including the Missouri Music Educators Association, are advocating for the reinstatement of the fine arts assessment.

The Show-Me Standards, Framework for Curriculum Development, and Assessment Annotations are available online at the DESE website:

[www.ese.state.mo.us/divimprove/curriculum/index.html](http://www.ese.state.mo.us/divimprove/curriculum/index.html)

*Deborah Fisher,  
Fine Arts Consultant, Missouri  
Department of Elementary &  
Secondary Education*

### **Pennsylvania**

The State of Pennsylvania has an Arts Assessment Sampler: Berks County Intermediate Unit 14 is the manager for the Arts Sampler project. Boyertown, Twin Valley and Philadelphia School Districts have done their work and created great designs for local assessment systems. They are trying a few assessments in their school district. Each Arts Assessment Sampler document went online summer of '01 via the PDE website. Workshops

and workbooks to manage the development of local arts assessment systems were to be made available throughout this academic year. The state is encouraging development of music assessments at the local level based on some shared models.

*Joan Rutkowski  
Penn State University*

---

*(continued from page 1 - Richardson)*

together a diverse group of music education leaders to present their own perspectives: Ed Marschlok of the New York State Department of Education, Dean James Undercoffler of the Eastman School of Music, and Dean Karen Wolff of the School of Music at the University of Michigan. The SRIG meeting will be held on Saturday, April 13 from 2:30 - 3:30 p.m. and I hope to see each of you there.

I urge each of you to take the initiative to find out what's happening on the music and arts education assessment front in your state, both in P-12 education as well as in teacher education, and send me a quick email report on what you find <richpete@umich.edu>. Future issues will contain updates from the other 44 states, along with important research initiatives within our field.

Once you know what is happening in your state, you might consider offering your professional expertise by becoming part of your state's team that wrestles with assessment issues. We are, after all, the experts on music teaching and learning, and we can make a difference, even in this politically charged climate.

---

*(continued from page 1 - Colwell)*

ment than in the use of inauthentic assessment. It has been rumored that some of our professionals have criticized contests and festivals as being "too similar" to the world of professional music.

A major role of this SRIG is to encour-

age research and development of the many possible evaluation devices-almost all research required indicators of success and efficiency. General music students need to know how well they are performing and what must be done to improve. Apparently the majority of Americans believe they have little "talent" for singing and that instruction would be of little value. The reader of *Education Week* recognizes that evaluation has been the most frequent story for the past several years and Russ Schultz, a dean of fine arts recently penned "Apples, Oranges, and Assessment" in *Arts Education Policy Review* indicating its importance at the tertiary level.

The field of measurement is rapidly moving forward using technology and sophisticated statistics but these advances are at present of little value to music educators as our first task is to build an audience for evaluation and to convince teachers, students, and administrators of its importance in teaching and learning. The 1997 national assessment created

less than a ripple among music teachers and no concern for change based upon any results. MENC's performance or opportunity to learn standards have had scant impact as MENC's resources have been devoted to disseminating information on the content standards, standards that are more subject to local values than are the performance or opportunity to learn standards.

Any state of school district that believes in a role for the arts standards finds itself having to find ways to determine the extent to which these standards are feasible and being met. Dr. Richardson has focused this effort that will require multiple revisions, revisions that can be facilitated by members of this SRIG and a sharing of item and program evaluation data. We are promised further enlightenment in future issues of this Newsletter. As least 13 states have accepted the idea of student assessment in the arts. Valid assessment strategies in music teaching are a concern not only PreK-12 but have permeated admittance to teacher

education, the effectiveness of in-service programs, exit exams, and are of concern in professional development. It has to be of interest to all readers of this newsletter that more emphasis is being placed on evaluating professional development than the conduct of a needs assessment of our present pre-service offerings.

#### Reference

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

Non-Profit Organization  
U.S. Postage Paid  
Ann Arbor, MI  
Permit No. 144

*Attention: Music Education Department*

University of Michigan  
School of Music  
1100 Bats Drive  
Ann Arbor, MI 48109-2085