Music
Educators
National
Conference

New England
Conference Of Music
290 Huntington Avenue
Boston, MA 02115

Number 17 Special Research Interest Group in Summer 1994

Measurement and Evaluation

Music Portfolios in Secondary Music Classes: What Do They Measure?

by Thomas W. Goolsby
University of Washington

Few movements in educational reform have grown so rapidly or attracted so much support from all quarters of American society as the idea of "Authentic Assessment." Further, probably no component of educational reform has been so blindly supported, misunderstood, or become subject to as broad an array of interpretation as the idea of "authentic assessment" (with the possible exception of "discovery learning" which made a brief, but significant appearance during the 1960's reform movement).

A review of scores of studies and reports completed since 1990 indicate that many writers use the terms authentic assessment, performance assessment, direct assessment, and alternative assessment interchangeably, while others point to specific characteristics of each. Alternative assessment is probably the most generic term which best reflects the rocketing support of an alternative means to assess public school student's achievement other than the traditional standardized, multiple-choice tests. All of these terms refer to some form of *direct examination* of student performance on tasks that are considered significant and relevant to life outside the school. While few advocates accede on the definition of terms, alternative assessment has received overwhelming support from politicians, public school administrators and policy makers, and thousands of teachers. Worthen reports that by the end of 1993, twenty-five states had enacted or were in the process of passing legislation *mandating* the use of alternative assessment to determine how individual schools and districts compared within the state. With Vermont and Connecticut leading the way, and Kentucky determined to have the first statewide completely performance-based assessment system by 1995, it would appear that politicians are accepting a new idea that looks good, but may not accomplish what is desired.[1]

The Use of Standardized Testing for "High-Stakes" Decisions

Possibly the source of strongest support from local and state politicians for alternative assessment has resulted from complaints about the use of standardized testing for "high-

New England Conservatory
Richard Colwell & John Holly, editors

stakes decisions." In addition to important decisions such as which students are promoted to the next grade, which students are placed in "advanced classes," or which students graduate, test scores frequently serve as the basis for important political and economic consequences such as where families purchase homes or which states are selected by foreign and domestic businesses for new factories. The emphasis on standardized testing and the pressure (both real and imagined) for teachers and school districts to maintain high test scores has culminated in the astonishing phenomenon that in 1988, *all* 50 states reported above average standardized test scores at the elementary level on at least one of the six major nationally-normed, commercial tests. [2] Again, advocates of alternative assessment point to these "high-stake" consequences as reasons teachers across the country have been pressured into "teaching to the test" — their response has been to create a "test" that is worthy of being "taught" — resulting in the development and use of the most popular form of alternative assessment: Portfolio Assessment.

SRIG newsletter #16 along with papers in this issue of SRIG #17, namely the preceding paper by Thomas Goolsby, and the paper to follow by Frank Abrahams, were presented at the National Convention of M.E.N.C., April of 1994, Cincinnati, Ohio.

Portfolio Assessment

It is also notable that currently many general education teachers who seem to support portfolio assessment do so not from a firm commitment to alternative assessment, but as a backlash to traditional standardized testing. Advocates and researchers in portfolio assessment recognize the problems they have encountered with reliability and refer to countless tomes in educational evaluation that remind teachers and researchers that validity is more important than reliability. [Validity referring to the idea that a test measures what it is *intended* to measure and reliability referring to the measure's *consistency*.] We all realize that it does not matter how reliable a measure is, if it fails to measure what it purports to measure, then it's not a good (or valid) measure.

So the questions is raised: *What do Music Portfolios measure* ? And the answer is not clear.

Many advocates of portfolio assessment would answer: *student growth*. And for that, portfolios probably are a valid measure; further, if the scale is known, perhaps they are even a reliable measure. But for assessing student growth, we don't need portfolios...we probably do not even need schools — we know it's going to happen and we know it's going to vary with each child. As Cizek has aptly stated regarding the movement toward alternative assessment: "we have begun a search for genuine-*looking*, authentic-*looking*, real-*looking* assessment and have eschewed more rigorous standards of validity." [3]

...countless tomes in educational evaluation...remind teachers and researchers that validity is more important than reliability.

...it does not matter how reliable a measure is, if it fails to measure what it purports to measure...

It would appear that much of the confusion surrounding portfolio assessment is that most politicians, administrators, policy-makers, and unfortunately teachers, do not recognize a difference between "assessment" and "evaluation." *Assessment* is normally used in reference to the means and tools used to gather information about students and their achievement while *evaluation* is used for the judgments and decisions educators make about students and their work using the results of the assessment.

Few secondary music teachers can be included in the thousands of teachers across the country who are fearful of, and under pressure from, the use of standardized tests for measuring student achievement. In fact, while virtually ignoring standardized music tests, music teachers have been involved in authentic assessment for generations. Each time a choral or instrumental music teacher rehearses, she is intensely engaged in assessing a multitude of performance elements (obviously performance-based assessment) and constantly conducting diagnostic evaluation (to determine *which* problems to address *when* and *how*). Clearly, in choir, band, and in orchestra, assessment guides instruction... and more effectively by some teachers than others.

Aspects of Portfolio Assessment

There are several aspects of portfolio assessment that should cause concern as well as a number of aspects that could prove beneficial to music education:

- First, on the negative side, while there does not appear to be any single "correct" way to design a portfolio, most writers advocate *a collection of students' work which includes examples of their "typical" work and their "best" work*. Now, perhaps this writer lives in a dream world, but one would like to imagine that we instill in our students a sense of achievement

where there is no difference between the two. Whether educators are referring to music or to general education for students who eventually will become assembly-line workers for General Motors, there should be concern for a national movement that recognizes, and perhaps rewards, students' growth in "typical" vs. "best" work.

• Second, most advocates of portfolio assessment want *the task of assessing student learning placed in the hands of teachers, based on the assumption that teachers are the most qualified to make informed decisions and judgments and are the best evaluators of their own students progress and achievements.* We all know from experience (and the research literature) that some teachers are less able than others. Abandoning objectivity and accountability easily could be more damaging than helpful. Further, accurate evaluation requires accurate assessment.

• Third, many advocates of portfolio assessment (and authentic assessment in general) are motivated to *abandon or alter the traditional grading system to enable more students to succeed (at their own pace). We in music education might be able to provide our colleagues in general education further insight into the* ***importance*** *given to a component of the curriculum where most students receive an A or a B.* For years Arts Advocacy groups have tried to convince our profession that music education might be taken more seriously if students were to receive the grades they deserved.

Characteristics of Portfolio

On the positive side, when viewed as a teaching tool rather than and end-all means for evaluation, there are several characteristics of portfolio assessment that could benefit music education. We have long been masters of ***diagnostic evaluation*** — as I mentioned, every time a choir or orchestra director rehearses, he is diagnostically evaluating the group and individuals. Gronlund and Linn consider diagnostic evaluation a "highly specialized procedure" where specific learning problems are identified and addressed through prescriptive instruction. [4] In music we normally call this rehearsal.

We are also proficient with ***placement and summative evaluation***: auditions, try-outs, concerts, festivals, and competitions. Placement evaluation is generally concerned with the student's background knowledge and initial performance level. Summative evaluation is analogous to the "final examination" and evaluates the quality of the final product or the level of achievement at the end of a unit, year, or course of study. [5]

Traditionally, however, we have neglected ***formative evaluation*** (that is, monitoring the learning of individuals to ensure that learning is taking place as opposed to simply identifying the blatant problems in a group setting). It is **this** area where the idea of portfolio assessment could benefit music education. The time spent in attending to individual learning might well be made up by less time required for performance preparation AND perhaps fewer individual students will be "lost in the crowd" (e.g., those students who spend their high school band career playing third clarinet).

Further, while few writers agree exactly on what should be included in a Portfolio, they all agree that it should include some type of *statement of purpose* written by the student and teacher together which addresses the goals and objectives of the semester or year. An **explicit purpose** prevents portfolios from becoming busy work and forces teachers to articulate goals and objectives — a purpose for doing what we do, a purpose for the curriculum and experiences that we arrange for our students.

Simply initiating a portfolio in music class forces the teacher to ask "*What is it that I want my students to learn and remember from this class?*" and this information is shared with the students. This idea is not new; we have known for years that students learn more when they are aware of what it is they are supposed to learn (in contrast to the music history exams where the "important" stuff is kept a secret until "finals week"). **Music education would be well served if music teachers articulated what it is that we want our students to remember when they leave our class.**

It would appear that much of the confusion surrounding portfolio assessment is that most politicians, administrators, policy-makers, and unfortunately teachers, do not recognize a difference between "assessment" and "evaluation." *Assessment is normally used in reference to the means and tools used to gather information about students and their achievement while evaluation is used for the judgments and decisions educators make about students and their work using the results of the assessment.*

Items Included in a Portfolio

The specific items that are included in a portfolio for secondary music naturally will depend on the purpose statement and the success of those who assume four essential roles:

(**a**) the ***designer***, usually the teacher (but in many school districts an administrator), yet in the most beneficial adoptions of portfolio assessment, a combined effort by teacher and student;
(**b**) the ***developer***, normally the student who decides what to include in the portfolio as evidence of learning achievement, not just a collection of finished

products, but evidence of progress;
(c) the ***evaluator***, as a basic concept behind portfolio assessment this responsibility would lie with the teacher (who makes the portfolios accessible to students, parents, and any other interested party); and finally, the least documented, and most ignored role:
(d) the ***conferee***, including both teacher and students in a one-on-one conference regarding the purpose statement, the collection of evidence, the progress toward the goals and objectives, and any amending of these to ameliorate the portfolio.

Authorities generally agree that portfolios should include something that the student finds difficult, something that the student has mastered, and something "in progress" (a work that the student is attempting to master or solve) — all of which are commonly found in a typical band student's folder. A fourth item included by many authorities include *attestations*, or documents *about* the student's work written by someone else. [6] This type of feedback is *rare* in a typical band folder. Even a simple checklist to document student achievements and progress would be of assistance to both teachers and students.

The model portfolio defined by Meyer, Schuman & Angello seems to include the major requirements that would satisfy all advocates. In brief, the portfolio should include a purpose statement, then a collection of evidence that demonstrates to the students (and any other interested party) the efforts and progress toward achieving the "purpose" including:
(a) student participation in what is collected for the portfolio's content;
(b) specific criteria for selection of the evidence;
(c) the criteria for judging merit — or the evaluation component;
(d) evidence of student self-reflection or self-evaluation. [7]

This and similar models have been met with much enthusiasm with the assumptions that such an assessment:

(1) will be viewed as using "real" examples of student classwork rather than isolated snippets of factual knowledge;
(2) the assessment will cover longer periods of time rather than a few hours; and
(3) the ideas that portfolios will enhance instruction.

... while virtually ignoring standardized music tests, music teachers have been involved in authentic assessment for generations. ***Each time a choral or instrumental music teacher rehearses, she is intensely engaged in assessing a multitude of performance elements...and constantly conducting diagnostic evaluation to determine*** *which* ***problems to address*** *when and how.*

Arts Propel

If we wish students to learn to listen critically, reflect on their own performance as well as the total ensemble performance (certainly a reasonable goal), then the Domain Projects developed by Arts Propel are viable tools and documents for inclusion in the portfolio since they have fulfilled the assumptions above.

Arts Propel was the first large-scale and systematic use of portfolio assessment in the arts. It was initially funded by the Rockefeller Foundation as a collaborative effort between the Educational Testing Service in Princeton, Harvard's Project Zero, and the Pittsburgh Public Schools. [8] Reports from this continuing program which has focused on portfolio assessment in music as well as the other arts have provided little data but, it's use has elicited praise from participants, researchers, and cautious observers.

Domain Projects

The Domain Projects are worksheets designed to be included in secondary music portfolios (e.g., a band folder) to facilitate student's self-evaluation of ensemble performance. The Domain Projects address production, perception, and reflection by requiring students to listen to taped performances of their own ensemble rehearsals and write comments evaluating the overall performance and their own personal performance. The process is repeated periodically, often using previously taped rehearsals for comparison and more in-depth reflection. Students learn to listen more critically, become aware of "more parts of the whole," and determine strategies to enhance their own performance and contribution to the total performance. [9]

Results of Systematic Studies

The results of the few systematic studies completed using portfolio assessment as measuring instruments for summative evaluation are not encouraging.

- Vermont's reliability coefficients for 1992 statewide portfolios range from .23 to .43 with 1.00 as the "strongest" relationship. [10]
- Baker and Linn report reliability as low as .16 for the Apple Class of Tomorrow Project in Ohio. [11]
- Hamp-Lyons and Condon found that the more information contained in Portfolios, the lower the reliability (which is the opposite effect in standardized tests where reliability is increased when more items are included on the test). They also discovered that different components of the portfolio were weighed differently. Evaluators usually reached a decision when evaluating the first piece of evidence and then used the remainder of the portfolio items to confirm their opinion or alter it only slightly. [12] This implies that if only the "showcase works" are evaluated, then the student with more parental support, or time, or any other facilitating assets will do better — which makes traditional standardized tests not only more

accurate, but less biased. On the other hand, if student growth or the learning *process* is to be evaluated, as most portfolio advocates emphasis, then the portfolios must include even more information to demonstrate this growth which will decrease the reliability even further.

• Further, cost becomes a significant factor with large-scale assessments. Reports from Vermont indicate that the previous computer-scored test cost about 10 cents per student. Colwell reports the cost of evaluating portfolios in Vermont rose to $37 per student in 1994 while the cost in larger states is slightly less (e.g., Texas reports $29/student). [13]

Returning to the initial question raised in this paper: *What do Music Portfolios measure*? remains problematic due to many issues that are raised during the introductory use of portfolio assessment across the country. Questions and issues such as:

• How long should an item remain in a portfolio?
• Should tapes of ensembles be included in individual's portfolios?
• Should teachers alter their curriculum to attain a better collection of documents in the portfolio?
• How do we balance the evaluation, for instance, of a pupil who joins choir late and progresses rapidly without attaining a high level of achievement with one who joins choir late but is already at such a high level that there is little progress?
• Do we evaluate the entire portfolio or just the final products (knowing that either way will affect the overall evaluation of individual students in opposite directions)?

• Further, one of the primary concerns of the pilot tests of using portfolio assessment in Rhode Island, Vermont, Kentucky, and other states has been the improvement of reliability of the measure through standardization — in effect, contradicting the basic purposes of utilizing alternative assessment in the first place. All of these questions and issues plus many more are still being argued by the advocates and writers in portfolio assessment.

It seems clear, however, that portfolio assessment may serve as a useful *teaching tool* if not a reliable or valid means for accurately measuring data for objective evaluation or for comparing states, school districts, or even individual students. Overall, the greatest benefit music education may glean from portfolio assessment is requiring teachers to clearly articulate their explicit purposes and goals for a semester year. A typical orchestra or choir folder may already contain much of what we may want to see in a music portfolio, but much is missing:

(**a**) written, documented feedback from teachers regarding individual and ensemble performance (with the overall goals considered);
(**b**) some type of checklist that sequences and documents each student's growth (including areas other than performance);
(**c**) written tests; and
(**d**) opportunity for reflection and self-evaluation such as the Domain Projects.

Also, while we have been well acquainted with authentic assessment for placement, diagnostic, and summative evaluation for many years, we have not done well in formative evaluation. Clearly some of our current assessment practices (primarily while on the podium) could be used for various types of evaluation, but the nature of portfolio assessment *does* get students more involved in the leaning/teaching/instructional process.

Footnotes

1 Worthen, B.W. (1993). Critical issues that will determine the future of alternative assessment, *Phi Delta Kappan*, 74 (6), 444-454.

2 The six tests include the *Iowa Test of Basic Skills, Stanford Achievement Test, Comprehensive Test of Basic Skills, California Achievement Test, Metropolitan Achievement Test*, and the *Science Research Associates*. Six states used locally constructed standardized tests and 18 states used one of the six commercial tests, but at the district level; Cannell surveyed the largest districts in these states. The six major tests allow for 90% of all school districts in the U.S. to report above average scores for the district and allow for 70% of all students to be told they are above the national norms for their grade level. This fascinating report is immediately followed by published letters of explanation by senior officials for the six testing companies. Cannell, J.J. (1988). Nationally normed elementary achievement testing in America's public schools: How all 50 states are above the national average, *Educational Measurement: Issues and Practice*, 7 (2), 5-9.

3 Cizek, G.J. (1991). Innovation or enervation, *Phi Delta Kappan*, 72 (9), 699.

4 Gronlund, N.E. * Linn, R.L. (1990). *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Company; 13.

5 *Ibid.*, pp. 12-13.

6 Barton, J & Collins, A. (1993). Portfolios in Teacher Education, *Journal of Teacher Education*, 44 (3), 200-210.

7 Meyer, C., Schuman, S. & Angello, N. (1990). *NWEA What Paper: Aggregating Portfolio Data*. Lake Oswego, OR: Northwest Evaluation Association.

8 Wolf, Dennie. (1987). What is arts PROPEL? *Portfolio*, 1 (1), 2

9 Davidson, L., Ross-Broadus, L., Charlton, J., Scripp, L. & Waanders, J. (1991). Music educators national conference 1990: recent advances in the state of assessment: Arts propel in Pittsburgh, *Measurement and Evaluation*. Colwell, R.J. & Ambrose, R. (eds.). Boston: Music Educators National Conference.

10 Kortetz, D., and others. (1992). The reliability scores from the 1992 Vermont portfolio assessment program. Interim Report. Rand Corp. Santa Monica, CA. ERIC document no. 355284.

11 Baker, E.L. & Linn, R.L. (1992) Writing portfolios: Potential for large scale assessment. Design and analysis of portfolio and performance measures. National; Center for Research on Evaluation, Standards, and Student Testing. Los Angeles, CA. ERIC document no. 350312.

12 Hamp-Lyons, L & Condon, W. (1993). Questioning assumptions about portfolio-based assessment, *College Composition and Communication*, 44 (2), 176-190.

13 Colwell, R.J. (1994). "Authentic assessment in music." Paper delivered to the Measurement and Evaluation Special Research Interest Group, Music Educators National Conference; Cincinnati, OH; April.

Portfolios in Student Teaching

by Frank Abrahams
Westminster Choir College of Rider University

Paper presented at the National Convention of M.E.N.C., April of 1994, Cincinnati, Ohio

This paper focuses on the practical application of portfolios as they impact the student teaching experience. Student teachers have always kept notebooks, made video tapes for evaluation purposes, and attended seminars with the cooperating teacher and the college supervisor. How are portfolios different? Do they enhance the quality of the student teaching experience, or are they fancy packaging for something we have always done?

Since portfolios measure process over time, it is logical that an appropriate test of their effectiveness would be in the student teaching experience. For three semesters, we have required students to maintain a portfolio during the student teaching semester at Westminster Choir College. Answers were sought to the following questions:

- Is the student teaching experience an appropriate venue for portfolios?
- Would grades be the same if we used the portfolios or graded traditionally?
- Would students come to know and understand the power of the portfolio model if they were to keep their own portfolios?
- How would students feel about the portfolio as opposed to the traditional grading format?

A set of student teaching outcome objectives was developed for each student to meet.

(See Figure 1.)

Have you changed your address since the last issue?

Figure I.

Westminster Choir College of Rider University
Department of Music Education

Student Teaching Outcomes

1. Initiate, plan and implement a unit of instruction which conforms to the Westminster criteria for a good lesson and the principles of "authentic learning."

2. Complete a lesson plan using:
   (a) the 4MAT System
   (b) Lesson Map
   (c) your co-op's planning format

3. Demonstrate your ability to:
   (a) engage students in cooperative learning
   (b) use the piano in teaching
   (c) use overheads or other "teacher made" visual aids
   (d) engage students in problem solving
   (e) engage students in higher level thinking
   (f) adapt instruction for children with special needs integrated into your class

4. Demonstrate computer competencies by:
   (a) maintaining an inventory on a database
   (b) maintaining a budget on a spreadsheet
   (c) preparing a concert program using Desktop Publishing
   (d) preparing a newsletter using Desktop Publishing

5. Attend a meeting of the:
   (a) Music department
   (b) School faculty
   (c) District school board

6. Teach two of the following:
   (a) a rote song
   (b) a listening lesson
   (c) an integrated arts lesson
   (d) an instrumental class, lesson or ensemble rehearsal

7. Select, rehearse, and perform in concert, a choral piece

8. Evaluate student performance and/or achievement by two of the following:
   (a) a short answer or true/false test
   (b) an essay test
   (c) a standardized achievement test
   (d) portfolio

The faculty were interested in testing the value of the student accepting the responsibility for collecting and then presenting evidence to support a particular grade.

Most powerful in the portfolio is the student journal. A list of the items that must be included is found in Figure V. The student must make a daily journal entry during Student Teaching. Journals are turned in each week at seminar, read by a faculty member, comments are written by that faculty member, and returned to the student. Lesson plans are also turned in and evaluated. This regular evaluation allows the faculty two advantages:

- The college supervisors are able to carefully monitor the student's teaching and to obtain insights into the student's thinking. Immediate intervention can be provided if necessary.

- The students are assured they can write anything and it will remain confidential.

The most difficult part of the portfolio assessment process was in developing the criteria by which the portfolio was evaluated. A five point likert scale was developed based on a model used by Educational Testing Services for the Praxis series of the National Teacher Examinations.

(See Figure II.)

Figure II

Westminster Choir College of Rider University
Department of Music Education

Student Name: _____ **Date**: _____

Teaching Assignment: Elementary Middle School High School **Placement** : _____

STUDENT PORTFOLIO EVALUATION FORM

Part One Teaching Outcomes

5 - 90% of the outcomes have been met and are superior in quality.
4 - 90% of the outcomes have been met and are satisfactory in quality.
3 - Most (75%) outcomes have been met and are satisfactory in quality.
2 - Most (75%) outcomes have been met; however, quality is uneven.
1 - Less than 75% of the outcomes have been met.

Part Two Lesson Plans

5 - Lesson plans are presented before every lesson taught or every rehearsed. All lessons meet the Westminster Criteria for a Successful Lesson
4 - Lesson plans are presented for every lesson taught or rehearsed. Most (75%) lessons meet the Westminster criteria for a Successful Lesson
3 - Lesson plans presented for every lesson taught or rehearsed. Lessons or rehearsals do not all meet the Westminster Criteria (i.e. inappropriate objective, do not reinforce high level thinking)
2 - Lesson plans are typed but do not exist for all lessons or rehearsals. Or, plans do not contain appropriate objectives and/or reinforce higher level thinking.
1 - Plans exist, but do not meet the Westminster criteria. Or, plans are missing for some lessons/rehearsals.

Part Three Personal Journal

5 - Journal includes specific instances where the student's thinking changed demonstrating the student's ability to analyze, synthe size and evaluate. Student has vision and the ability to dream.
4 - Personal journal contains insights and clearly demonstrates the student's ability to analyze and synthesize. Student has vision and the ability to dream.
3 - Journal includes specific instances where the student's thinking changed but does not demonstrate the student's ability to solve problems or think at higher levels (analysis, synthesis, evaluation). Student shows no vision and does not have the ability to dream.
2 - Journal is "gossipy" and does not provide evidence of reflective teaching.
1 - Journal entries are missing

Part Four Observation Reports

Total scores for each observation are to be converted to a scale from 5 to 1 and then an average number given.
Total Scores _____ Converted Scores (5-1) _____
Total Score _____ Letter Grade _____ 5 = A • 4 = B • 3 = C • 2 = D • 1 = F © FA, 1993

Students were asked to complete a release form allowing the faculty to use the portfolio for various purposes.
(See Figure III)

Figure III

Westminster Choir College of Rider University
Department of Music Education

Portfolio Release Form

I give my permission for the music education department at Westminster Choir College to use the contents of my portfolio for the purposes of demonstration, educational workshops, seminars, clinics, classes, publications and research. I understand that my name will not be identified, nor will the name of the school or cooperating teacher be identified.

_____ YES _____ NO

I give permission for the music education department at Westminster Choir College to show my video tape for teaching purposes, demonstration or to use it in research. I understand that every effort will be made not to identify me by name.

_____ YES _____ NO

_____ _____
signature date

_____
print name

Lessons and lesson plans were measured against a criteria in addition to their weight in the portfolio.
(See Figure IV)

Figure IV

Westminster Choir College of Rider University
Department of Music Education

Criteria for a Successful Lesson

1. The lesson is centered around a musical concept(s) that engages problem solving.
2. The teacher combines strategies which engage the right brain as well as the left.
3. There is music or music making in the lesson.
4. There are strategies which address the diversity of learning types.
5. Children are asked probing questions that motivate higher order thinking and that involve children in using musical thinking to solve the problem.
6. Music activities and tasks are authentic.
7. The teacher checks for understanding at multiple points within the lesson using a variety of evaluative tools.
8. The activities provided in the lesson present the evidence students need to solve the problem and provide the preparation for the next music learning experience.
9. Musical skills, including but not limited to "audiation" are being developed.
10. The lesson content is aesthetically sound.

The entire portfolio consisted of seven sections.

Figure V

Westminster Choir College of Rider University
Department of Music Education

Name _____

PORTFOLIO INVENTORY

Part One: Teaching Outcomes _____

_____ 1. A unit of instruction initiated by the student that conforms to the Westminster criteria for a good lesson/ rehearsal and demonstrates the principles of "authentic learning."

2. A lesson plan using:
_____ a. the 4MAT System
_____ b. Lesson Map
_____ c. the co-op's planning format

3. The ability to:
_____ a. engage students in cooperative learning
_____ b. use the piano in teaching
_____ c. use overheads or other "teacher made" visual aids
_____ d. engage students in problem solving
_____ e. engage students in higher level thinking
_____ f. adapt instruction for children with special needs integrated into your class

4. Computer competencies by:
_____ a. maintaining an inventory on a database
_____ b. maintaining a budget on a spreadsheet
_____ c. preparing a concert program using Desktop Publishing
_____ d. preparing a newsletter using Desktop Publishing

5. Student's attendance at a meeting of the:
_____ a. Music Department
_____ b. School faculty
_____ c. District School Board

6. The student's teaching two of the following:
_____ a. a rote song
_____ b. a listening lesson
_____ c. an integrated arts lesson
_____ d. an instrumental class, lesson or ensemble rehearsal

_____ 7. The ability to select, rehearse, and perform in concert, a choral piece

8. The ability to evaluate student performance and/or achievement by two of the following:
_____ a. a short answer or true/false test
_____ b. an essay test
_____ c. a standardized achievement test
_____ d. portfolio

Part Two: Lesson Plans _____

Part Three: Personal Journal _____

Part Four: Observation Reports _____

_____ a. Observation 1 Grade _____
_____ b. Observation 2 Grade _____
_____ c. Observation 3 Grade _____
_____ d. Observation 4 Grade _____

Part Five: Seminar Assignments _____

_____ a. Five Year Projection
Grade _____
_____ b. Dream Budget
Grade _____
_____ c. Job Application Questions
Grade _____
_____ d. Philosophy of Music Education
Grade _____

Part Six: Video Tapes

_____ a. Week 3 Grade _____
_____ b. Week 6 Grade _____
_____ c. Week 9 Grade _____

Part Seven: Evaluation Forms

_____ a. From the Cooperating Teacher
_____ b. Of the Cooperating Teacher

Total Score: _____, **Grade:** _____,

Grade Conversion: A = 5 • B = 4 • C = 3 • D = 2 • F = 1

At the conclusion of the practicum block, each student attends a portfolio review with the members of the department. At the review, the student presents the portfolio. A video of their best teaching, is shown, the best lesson plan examined, and readings from the journal where a problem was solved, thinking changed, and an example of the student's ability to think critically, synthesize and evaluate are presented. Core values and a philosophy of music education are discussed. Each faculty member records commendations and suggestions for improvement, and a grade for the portfolio review session. The comments are collated and the student receives a summary copy with a duplicate copy placed in the student's file. The composite portfolio grade is averaged into the seminar grade. The college supervisor gives a separate grade for the practicum experience itself.

(See Figures VI and VII page 10)

Figure VI.

Westminster Choir College of Rider University
Department of Music Education

Student Name _____

PORTFOLIO REVIEW EVALUATION FORM

Commendations:

Areas identified as weak or in need of improvement:

Recommendations:

Evaluator: _____ **Date:** _____

Figure VII.

Westminster Choir College of Rider University
Department of Music Education

PORTFOLIO REVIEW INSTRUCTION SHEET

1. Place the portfolio in a loose leaf binder and arrange the contents in the order of the inventory sheet. Use separators for each major part of the portfolio (**PART ONE, TWO, THREE, FOUR, FIVE, SIX, SEVEN, EIGHT**). Your video tape(s) may be included separately at the end. Include the inventory sheet separately.

2. Check off on the inventory sheet those items you have included.

3. Identify each item in **PART ONE** by number in the upper right hand corner of the item.

4. Arrange the lesson plans chronologically in **PART TWO.** Do not retype any plans as they must include all comments made over the Practicum.

5. Put a Post-it-Note (forming a TAB) on your best plan. Indicate on the Post-it-Note why it was chosen. Be prepared to speak about it at the Portfolio Review.

6. Do not retype your journal. It must include all of the comments made over the course of the Practicum. Put a Post-it Note (forming a TAB) at an entry that:
   (a) shows your ability to solve a problem
   (b) a change in your thinking
   (c) your ability to analyze, synthesize, and evaluate
   (d) your best entry

On the Post-it-Note explain why each was chosen. Be prepared to speak about them at the Portfolio Review.

7. Cue your video tape to show your best teaching. Be prepared to discuss your tape.

8. Dress professionally for your portfolio review.

The results of the three semester study are not finalized. The students appear to enjoy the portfolio format. They like the opportunity to write in the journal, and their like the immediate feedback and tight monitoring and they being a partner in the portfolio review process. The students are probably receiving the same grades as they would if the the process were traditional; however, they are thinking more critically and becoming more thoughtful about their teaching.

Future issues of this SRIG...

will document evaluation schemata being developed by states and the preparation for the 1996 national assessment program.

Your editors welcome the submission of manuscripts for inclusion in the SRIG newsletter.
Also welcome are news items about assessment and evaluation activities in your area.

NAEP to CCSSO to ETS to the Public: Designs for the 1996 National Assessment of Arts Education

by: Thomas W. Goolsby, University of Washington

While 1996 may seem distant to those who favor "Leap Years"....or to the GOP.....To those involved in developing and implementing the next National Assessment of Educational Progress, 1996 is rapidly advancing. "The Nation's Report Card," the National Assessment of Educational Progress (NAEP) is a congressionally mandated project of the National Center for Education Statistics under the U.S. Department of Education. The NAEP, however, has been in existence since 1969 (before the Department of Education existed) and conducted the first nationwide assessments as administered by the Education Commission of the States and funded by the National Institute of Education. In 1988 Congress created the National Assessment Governing Board (NAGB) to formulate policy guidelines for NAEP. The board assumes responsibility for determining what subject areas of American schooling will be assessed, including the authority to add to those subject areas specified by Congress. The 24 members of the NAGB are delegated the responsibility to:

- "identify appropriate achievement goals for each age and grade;
- develop assessment objectives;
- develop test specifications;
- design the assessment methodology;
- develop guidelines and standards for data analysis and for reporting and disseminating the results;
- develop standards and procedures for interstate, regional and national comparisons;
- improve the form and use of the National Assessment; and...
- ensure that all items selected for use in the National Assessment are free from racial, cultural, gender or regional bias" (College Board, 1994, p. iv).

Music was included in the 1972 and 1978 assessments (visual arts in 1975 and 1978), and will once more be assessed in 1996. In January of 1992, the NAGB Governing Board solicited proposals to develop an assessment framework and specifications for the 1996 Arts Education Assessment. The contract was granted to the Council of Chief State School Officers (CCSSO) who then subcontracted The College Board and the Council for Basic Education to assist in developing the Assessment Framework and the Educational Testing Service (ETS) for the actual test construction. Together these groups worked on the "consensus project" involving a multitude of representatives from Federal and State governments, professional organizations, and educational groups. Table 1 (p.16) contains the National Initiatives in a timeline. Table 2 (p.16) provides an outline of the major products of the "consensus project."

In 1988 Congress created the National Assessment Governing Board, NAGB, to formulate policy guidelines for the National Assessment of Educational Progress, NAEP. *The board assumes responsibility for determining what subject areas of American schooling will be assessed...,* ***Music was included in the 1972 and 1978 assessments, (visual arts in 1975 and 1978), and will once more be assessed in 1996.***

The CCSSO initiated its task by appointing a 32-member Planning Committee (chaired by Frank Phillips) to develop and recommend not only the assessment framework, but other design features for an assessment to include dance, music, visual arts, and theatre. This aspect of the overall, enormous project was guided by a 29-member Steering Committee including an even broader representation of American Society than the Planning Committee. The development of the Assessment Framework started in September of 1992 and was concluded 18 months later in March 1994.

Simultaneously, the *National Standards for Arts Education* was being developed by the Consortium of National Arts Education Association (comprised of representatives from the American Alliance for Theatre & Education, Music Educators National Conference, National Art Education Association, National Dance Association). Realizing in 1992, the rare opportunity for a National Assessment in the Arts coinciding with the rapidly developing standards project, the NAGB appointed several individuals to leadership positions on both projects (e.g., Graham Down, Co-Chair of the CCSSO Steering Committee of the assessment project was named Chair of the Oversight Committee for the Consortium's Standards project). "This confluence of a standards-setting process and its immediate application in creating a national assessment provide an unprecedented opportunity to align standards and assessment in a model for arts education" (NAGB, 1994, p.2).

NAEP Framework

The NAEP has organized the arts in a slightly different format than the published *National Standards for Arts Education* (MENC, 1994). NAEP has structured the Framework through *processes* and *content.* Processes include: creating, performing, and responding. Content is comprised of two major components:

(1) knowledge and understanding about the arts,
(2) perceptual, technical, expressive, and intellectual/reflective skills.

Through the processes of responding, performing, and creating music, students draw upon and apply this knowledge and these skills.

Table 5 contains the NAGB Framework Matrix for the 1996 NAEP. (see p.19) The diagram illustrates each of the art discipline's approaches to the common framework. Each cell represents a sub-scale where results *may* be reported (emphasis in the document; NAGB, p.20). Table 3 (see p.17) outlines specifically how this Framework is applied to music: the results of the overall consensus of how to assess knowledge and skills in music.

Each of the three processes described in Table 3 will be assessed at the 4th, 8th, and 12th grade levels. Data will be reported in nominal categories: basic, proficient, and advanced.

- Basic denotes partial mastery of the content but performance that is only fundamentally adequate at the three grade levels.
- Proficient represents solid academic achievement and competency over challenging subject matter.
- Advanced performance on the assessment represents achievement that is equal to that expected of top students (NAGB, p.51).

The Framework concludes with the content outlines for each of the four arts for each grade level to be assessed. In the case of music, The NAEP Planning Committee used the *National Standards for Arts Education* developed by the Consortium of National Arts Education Association; at this stage the nine standards in music along the three grade bands (MENC, 1994) were converted to the three artistic processes to be assessed. The NAEP modified the wording of the Standards in order to provide the level of detail needed by assessment developers, that is, specific *tasks* that would demonstrate mastery of the standards. Throughout the content outlines included in the Framework, these additions appear in italics — the resulting outline is better organized, with greater clarity than the document published by MENC.

Another aspect of the Framework (NAGB, 1994) that makes the 1996 NAEP Arts Assessment different from the previous music assessments, was the task of the Steering Committee "to balance 'what is' in U.S. arts education with 'what ought to be' " (p. 5). Previous NAEP music assessments have focused on "cognitive abilities in music, some of their attitudes about music and some of their experiences with music. The second assessment included exercises from the first in order to determine changes in these areas over time" (NAEP, 1981, p. 1). Both the first and second music assessments utilized test booklets (packaged with social studies exercises during the 1971-72 assessment and writing exercises during the 1978-79 assessment). The 1996 assessment will measure music through performing, responding, and creating.

Further, by combining efforts of the 1996 NAEP with the *National Standards for Arts Education*, a genuine attempt is being made to determine what students are learning in music education. Reports from the previous music assessments recognized the difficulties in assessing achievement in music due to the "fact that many of the primary goals of music education cannot be easily stated in terms of observable behavior"...and regarding the second music assessment: those involved in developing the new objectives suggested that data be gathered on the musical training background of each respondent" (NAEP, 1981, pp. 3-4). The result of the 1996 NAEP in music is directly focused on what is being accomplished in American music education. The College Board, however, has advised the NAGB to recommend to ETS that "whenever exercises from the previous NAEP music assessments appear valid and address the content outline [ETS] should consider making use of those exercises to permit comparisons between present students' achievement and that of students in earlier NAEP music assessment" (College Board, 1994, p. 30). By the third meeting of test-writing teams (June 1994), exercise developers were verbally told that none of the previous items would be used.

In January of 1992, the NAGB governing board solicited proposals to develop an assessment framework and specifications for the 1996 Arts Education Assessment...*The development of the assessment framework started in September of 1992 and was concluded in March 1994....Simultaneously, the* ***National Standards for Arts Education*** *was being developed by the Consortium of National Arts Education Association...,* ***"This confluence of a standards-setting process and its immediate application in creating a national assessment provide an unprecedented opportunity to align standards and assessment in a model for arts education"*** *(NAGB, 1994, p.2.)*

Test Specifications & Guidelines

As a result of current emphasis on authentic assessment, an attempt is being made to implement methods of alternative assessment to the 1996 NAEP; in brief, unlike the two previous music assessments, very few multiple-choice items will be used and emphasis will be placed on construction-response items. Ten music teams (each comprised of four to six music educators from a single state) across the country have been asked to develop integrated test exercises for the NAEP. The first two teams met on 14-16 April in Portland, OR. comprised of music educators from Washington and Oregon. The three day meeting was led by Frank Phillips, Consensus Coordinator for the CCSSO, who immediately started a 40-day non-stop tour of the country in coordinating and instructing the various music teams around the country. [Fifteen states are involved in test construction; each state has been asked to establish two teams in separate arts disciplines; ten states have been working in music.]

The overall plan is to generate a minimum of 180 integrated exercises (for music), each consisting of multiple scorable units, and select the most valid and reliable exercises for the 1996 NAEP. The music portion will be limited to 60 minutes for 4th graders and 90 minutes for 8th and 12th graders. Students in the national sample will be assessed in a single arts area. Students within the same school may even receive different exercises within a single arts area in an attempt to measure performing, responding, and creating within the 60-90 minute time frame. The test items not selected by NAEP will be made available by the CCSSO for state assessments.

The NAEP Arts Assessment will include exercises in three formats: (1) production exercises, (2) open-ended questions requiring written responses, and (3) multiple choice items. The first two types of exercises are labeled constructed-response exercises and will comprise 80-90% of each integrated exercise. Multiple choice items are limited to 10% of the total exercise. For the exercise developers, this becomes quite problematic in that the test will be administered in three "blocks:"

- Block A *is designed to measure "breadth" of knowledge* and will be administered to intact classes of students.
- Block B *is designed to measure "depth" of knowledge* and will also be administered to intact classes of students at the three grade levels. Differences between these two blocks are in the assessment exercises to be administered; both types of assessment tasks will be constructed in test booklets.
- Block C *is comprised of a "subsample"* which will be selected from various schools and designed to administer individual exercises.

It is obviously Block C for which many or most of the performance exercises are designed, and the NAGB Framework indicates that the music exercises be allocated 30-40% performing, 25-45% responding to music and 20-30% for creating. Considering the large portion of the standards that require performance (including, for example, improvisation, which falls under the category "creating"), it would appear that relatively few students will be included in the major categories, that is, the subsample or Block C. These and other limitations are what faced the test developers who were given a deadline of 25 June to complete all items to be forwarded to ETS.

As a result of current emphasis on authentic assessment, an attempt is being made to implement methods of alternative assessment to the 1996 NAEP...*very few multiple-choice items will be used and emphasis will be placed on construction-response items.* ***Ten music teams comprised of 4-6 music educators from a single state across the country have been asked to develop integrated test exercises for the NAEP...****Fifteen states are involved in test construction...ten states have been working in music.*

Integrated Exercises

The design of the Arts Assessment is based on *integrated exercise sets*. These include a series of related exercises in various formats that are built around common stimulus material (e.g., a brief musical recording). Each related exercise may be comprised of several scorable units. This approach is justified by the limited time for the assessment and belief that students will not have to mentally "shift gears" (ETS, 1994, p. 5). An example might be playing a recording (instructions also recorded) and asking 4th graders to draw numbers on a world map to indicate where each musical example originates; the recording might then present musical excerpts from the Baroque period, Java, Africa, Asia, Jazz, and so on. This single exercise would assess breadth and fall in the Block A category. Subsequent exercises about these same musical examples (e.g., what they have in common, how do they differ, etc.) help complete the integrated set and provide a measure of depth — Block B. Then a subsample of student's would be selected and asked to individually improvise (or compose) a rhythmic selection based on one of the original stimuli that was measured in depth. Through this "integrated exercise," various tasks would be assessed and scored. An additional dilemma facing exercise developers is that several of the production exercises are to be completed/performed in groups. While this clearly is the way much music is taught and performed, the guidelines specify that the develop-

ers must devise ways for individually scoring students or members of the group.

Some assessment developers have been quite successful in "scaffolding" integrated exercises. The final single exercise is quite complex "on paper" and in description, but works well in assessing breadth, depth, and individual constructed-responses based on a single "idea" or on a single initial idea. Students are provided directions and stimulus for the first exercise which may be multiple choice questions to measure cultural or historical context (and to get the students "thinking in the right direction"). When this scorable unit is completed, the same stimulus is presented with a different set of instructions to encourage students to write responses to open-ended questions that require describing music in musical terms. This exercise may then be followed by the same stimuli and a similar one to solicit evaluation or reflection of musical performances. Finally a subsample is selected to improvise in the style that has now been heard several times. Such integrated exercises do not overwhelm children with a multitude of tasks, but guides them through tasks representing three to five standards at various levels of achievement.

Pilot Testing & Sampling

Following the task of developing integrated exercises by the 10 state music teams during the months of April, May, and June, the items were collected by ETS. An initial computer data base was created to establish a frequency distribution of the number of scorable units for each content standard for each proficiency level for each grade level. ETS will appoint an "assessment development panel" which will review each set of scorable units (crossed by integrated exercises) to determine grade-level appropriateness, content validity, development cost, and administration cost/efficiency. For example, the data base will list all exercises that include the 8th grade scorable unit: "RESPONDING, B.2. identify and define (*while looking at a score*) standard notation symbols for pitch, rhythm, articulation (*accents, legato, staccato, marcato*), dynamics (*piano, forte, crescendo, diminuendo*), tempo, and expression (*phrasing*)" (italics added to the *National Standards* document by The College Board, 1994, p. 131). Several of these items may have been written as stand-alone items, while others may be part of integrated exercises that also assess other standards. The "assessment development panel" will look at content validity, cultural, and gender bias, and the efficiency of reliably assessing the maximum number of standards within a single exercise (i.e. cost efficiency, time efficiency, and usability). Possible results may include revisions such as replacing specific scorable units by one exercise author with more valid and more efficient scorable units by another author.

The E.T.S. reviewers will also maintain the original integrity of the exercise pool balanced across the three artistic processes (performing, creating, and responding) and the two learning components (knowing about and doing), and across grade levels. Their primary task is to select and combine the individual scorable units extracted from the integrated exercises to create an assessment instrument that has validity (i.e., measures as many of the Standards as possible) and is efficient in terms of both money and time.

The selected exercises will be pilot tested by a field testing firm (the organization was not confirmed by 27 June 1994). The purpose of the pilot test will be to determine the "usability" of various integrated exercises (e.g., space, resources, audio and video accessibility and reliability, and electronic and acoustic musical instruments). As in previous NAEP testing, recording devices will operate continuously during the pilot testing to record student questions and problems in understanding expectations. These data are then used to drop or refine some test items (e.g., instructions). Also determined during the pilot tests will be an accurate measure of the time required to administer various exercises. While exercise developers have been required to indicate the length of time necessary for each scorable unit, these estimates may prove unreasonable (e.g., the Block C items that are administered to individuals). Other aspects involve determining, for example, if dance can accurately be assessed on videotape or will it require "on-the-spot scoring." An analogous problem holds true for music items assessing the Standards associated with playing in an ensemble and/or following a conductor. Finally, the results of the pilot test will also establish the extent that the test administrators need to be trained musicians.

The sample for the pilot test will be drawn from the same multi-stage stratified sampling used in previous NAEP assessments: grade level and demographic subgroups. The pilot tests are initially scheduled to be given during the middle of the 1994-95 school year. These tests will be conducted in several of the 15 states involved in the test construction (CA, CT, GA, FL, IL, IN, MD, NC, NY, OR, PA, TX, WI, VT, and WA). The 1978 NAEP Music Assessment sampled in such a way that approximately 2500 students responded to each item. Until the pilot testing is complete and the cost of administration and scoring is factored in, the 1996 sample size will not be determined. With at least 120 separate standards listed in the Music Content Outline (College Board, p. 126-135), and imagining that creative exercise developers can assess 10 standards in one or two integrated exercises within 60-90 minutes, a total sample of 30,000 would be required for just the music portion of the Arts Assessment to equal the number of respondents per item on the 1978 Assessment. This figure exceeds by about 15% the number discussed for the total sample for the 1996 Assessment for all four arts areas.

An additional subsample has been identified through the Planning Board's desire to assess music in exemplary programs. Preparatory Schools of Music as well as public school districts that offer all four arts in grades K-12 have been targeted for this separate sample.

Scoring

The scoring rubrics for integrated exercises are extremely difficult to develop and describe. They might be considered a psychmetricians dream and a graduate Test & Measurement student's nightmare. The idea is to glean as much information from each scorable unit as possible. Many of the open-ended questions or production exercises lend themselves quite readily to assessment of numerous simultaneous tasks, but the actual scoring of such constructs such as tone quality (proficient for 4th grade, 8th, grade, and 12th grade, for example) and intonation requires either an inordinate amount of training, or the use of well-trained musicians, or high technology. The latter has been mentioned in several meetings as a means to score production activities in dance and the visual arts.

The idea is to glean as much information from each scorable unit as possible.

...since process is such an integral part of the Standards, much of the scoring must attempt to measure process in lieu of a finished product. Many of the scoring rubrics do not utilize a "right and wrong" measure, but attempt to determine quality of a process...a very difficult task indeed.

Further, since process is such an integral part of the standards, much of the scoring must attempt to measure process in lieu of a finished product. Many of the scoring rubrics do not utilize a "right and wrong" measure, but attempt to determine quality of a process.....a very difficult task indeed. For example, production exercises "require students to carry out artistic processes they have practiced (i.e., "doing music). These exercises should closely parallel important lifelong artistic behaviors and should be as faithful as possible to artistic learning. The exercises and the students' responses to the exercises need to reflect the way real artists think and work" (College Board, p. 84). Such exercises may be developed, but the scoring becomes problematic. Assessing improvisation has baffled many of the finest researchers in music education for decades — in the NAEP it is compounded by requiring improvisation of a variety of cultural styles and genres, plus assessing the student's personal evaluation of the product. All of this ostensibly based on the "way real artists think and work."

Additional scoring problems arise due to the hierarchical nature of the Content Outline. There will undoubtedly be many fine young musicians who do not have the verbal facility to describe or discuss what they know; lack of appropriate musical terminology will result in "basic" scores rather than "proficient" or "advanced" scores. At the fourth grade level of Creating, "use a variety of sound sources when composing" is two levels above "create music to accompany readings or dramatizations, *manipulating dimensions such as the variety of sounds, tempo, loudness, and mood of piece to enhance or match the readings or dramatizations, and describing and explaining the choices made*" (College Board, p. 126). This latter standard contains eight separate tasks (scorable units) which are obviously not equal but are ipsative in nature.

Summary

There is a remarkable amount of idealism driving the 1996 NAEP Arts Assessment. Many of the leaders on the Steering and Planning Committees visualize what children in the United States should know and be able to do. Many of the actual test developers cannot help being affected by what they realize their own 4th or 8th grade students know and can accomplish. What is being developed for the 12th grade students will be administered only to 12th graders who have been in specialized music courses throughout high school (i.e., band, choir, and orchestra or private instruction). The remaining 12th graders will be administered the 8th grade assessment in order to determine effects of specialized electives or training in music. This decision in itself indicates that the Committees are beginning to take a more realistic view toward arts education in America.

Table 4 indicates the timeline that is being followed by the various committees and subcommittees working on the project. Without a doubt no previous NAEP has had so much input or been so ambitious as what is planned for 1996. The results may indicate much about arts education's current status; many of us can already anticipate what many of the results will reveal.

References

College Board, The (1994). *Arts Education Assessment and Exercise Specifications* (Pre-publication edition). Approved by the National Assessment Governing Board; 5 March 1994.

Educational Testing Service (in collaboration with the Council for Chief State School Officers) (1994). *Guidelines for NAEP Arts Assessment Exercise Developers*. Princeton, NJ: ETS.

Music Educators National Conference (1994). *National Standards for Arts Education*. Reston, VA: MENC.

National Assessment of Educational Progress (1981). *Music 1971-79: Results from the Second National Music Assessment*. Denver: Education Commission of the States.

National Assessment Governing Board (1994). *Arts Education Assessment Framework* (Pre-publication edition). Prepared by the Council of Chief State School Officers with The College Board and The Council for Basic Education. 5 March 1994.

Table 1

National Initiatives in Context

Nation at Risk
1983

National Goals
1989

National Council of State Arts Education Consultants
1991

National Council for Standards and Testing (NCEST)
1992

National Standards for Arts Education
1992-1994

Consensus Project for a National Assessment (NAEP)
1992-1994

The Arts as Part of the GOALS 2000 Legislation
1994

Table 2

The Major Products of the 1996 NAEP Consensus Project

The Assessment Framework:

This document provides broad descriptions of what the assessment should be like. It is directed at a general audience, yet has enough specific detail and content specificity to be a useful tool for teachers and assessment specialists.

The Assessment and Exercise Specifications:

This document is like the blueprint for the design, providing the specific details for the test development contractor on how to construct the assessment. Basically a technical description, this version also gives answers to psychometric questions and defines the parameters and methods for scoring the assessment.

The Reporting Strategies:

The Reporting Strategies Document answers questions about what types of information are important to whom and suggests ways of reporting that information in a timely and effective manner.

The Background Questions:

The Background Questions produce significant information about the educational environment of the students participating in the national sample. This material is necessary to fairly and wisely interpret the assessment results in the context of the school's resources and the student's opportunity.

Table 3

Music Assessment Framework

Arts Processes For Music

Creating - When improvising, composing, or arranging music, students:

- apply historical, cultural, and aesthetic understanding by creating stylistically appropriate alteration, variations, and improvisations;
- use standard and/or non-standard notation to express original ideas;
- evaluate, refine, and revise successive versions of original work;
- demonstrate skill and expressiveness in the choice and use of musical elements; and
- present the created work for others.

Performing - When singing or playing music with musical instruments, students:

- select appropriate repertoire;
- apply skill by performing with technical accuracy;
- develop an appropriate and expressive interpretation by applying understanding of structure and cultural and historical contexts of music;
- read musical notation accurately;
- evaluate, refine, and revise the performance; and
- present the performance for others.

Responding - When perceiving, analyzing, interpreting, critiquing, and judging, music students:

- select repertoire for listening;
- analyze the elements and structure of music;
- compare and contrast various musical styles;
- identify formal and expressive qualities that distinguish a particular style of music;
- place music within its cultural and historical context;
- make critical judgments about technical and expressive qualities of musical performances and compositions; and
- use movement or words to interpret and describe personal responses to music.

Based On Specific Content In Music

Knowledge	**Skills**
Applying Knowledge of:	Applying Cognitive, Affective, and Motor Skills Including:
Context: • Personal • Social • Cultural • Historical Aesthetics Form and Structure Processes	• Perceptual • Intellectual/Reflective • Expressive • Technical

Table 4

Timeline

NAGB Action On Framework and Specifications
March 1994

Test Development Begins
March 1994

NAGB/NCES Review of Field Test Exercises
Late Summer 1994

Revisions and Field Test Preparations
Fall 1994

Field Test Administered
February 1995

Field Test Responses Scored and Analyzed
Winter/Spring 1995

Development of Operational Assessment
NAGB Review of Final Exercises
Summer 1995

Revisions and Operational Assessment Preparation
National Assessment Administered
February 1996

Responses Scored and Analyzed
Spring/Summer 1996

Results Released
1997

Teacher Portfolio Assessment

by Peter Doolittle
The Catholic University of America

If you are a teacher, are you a good teacher? Would you like to be a better teacher? If you are an administrator, are the teachers for whom you are responsible doing a good job? Assessing the productivity, efficiency and effectiveness of teachers is a formidable task. While the National Teacher's Exam may provide a minimum criterion for the certification of teachers, it is not meant to be used as a measure of teacher effectiveness. One method for assessing teacher performance is the teacher portfolio.

What Is a Teacher Portfolio?

A teacher portfolio is a collection of work produced by a teacher. Just as an artist uses a portfolio of collected works to illustrate his or her talents, a teacher portfolio is designed to demonstrate the teacher's talents. Thus, teacher portfolios are constructed by teachers to highlight and demonstrate their knowledge and skills in teaching. A portfolio also provides a means for reflection; it offers the opportunity for critiquing one's work and evaluating the effectiveness of lessons or interpersonal interactions with students or peers.

What is actually included or related in a teacher portfolio depends on how the portfolio will be used. A portfolio may include some or all of the following:

- Teacher background.
- Class description: time, grade, and content.
- Written examination: National Teacher's Exam, State licensure tests.
- A personal statement of teaching philosophy and goals.
- Documentation of effort to improve one's teaching: seminars, programs, etc.
- Implemented lesson plans, handouts and notes.
- Graded student work such as tests, quizzes and class projects.
- Video/audio tape of classroom lessons.
- Colleague observation records.
- Written reflections on teaching.
- Photographs of bulletin boards, chalkboards or projects.

A common misconception is that a teacher portfolio is a folder laden with teaching artifacts and evaluations. Ideally, a teacher portfolio is a document created by the teacher that reveals, relates, and describes the teacher's duties, expertise, and growth in teaching. Each assertion in the portfolio is then documented in an appendix or a reference to outside material, such as videotapes or lengthy interviews. The size of a portfolio varies, but it is typically two to ten pages, plus appendices.

How Is a Teacher Portfolio Used?

A teacher portfolio is an education tool, which is primarily used in two ways. First, portfolios are used as a means of authentic assessment in evaluating the effectiveness of a teacher for licensure and/or employment decisions. Second, teacher portfolios are used to provide feedback to teachers so that they may improve their teaching and level of professionalism.

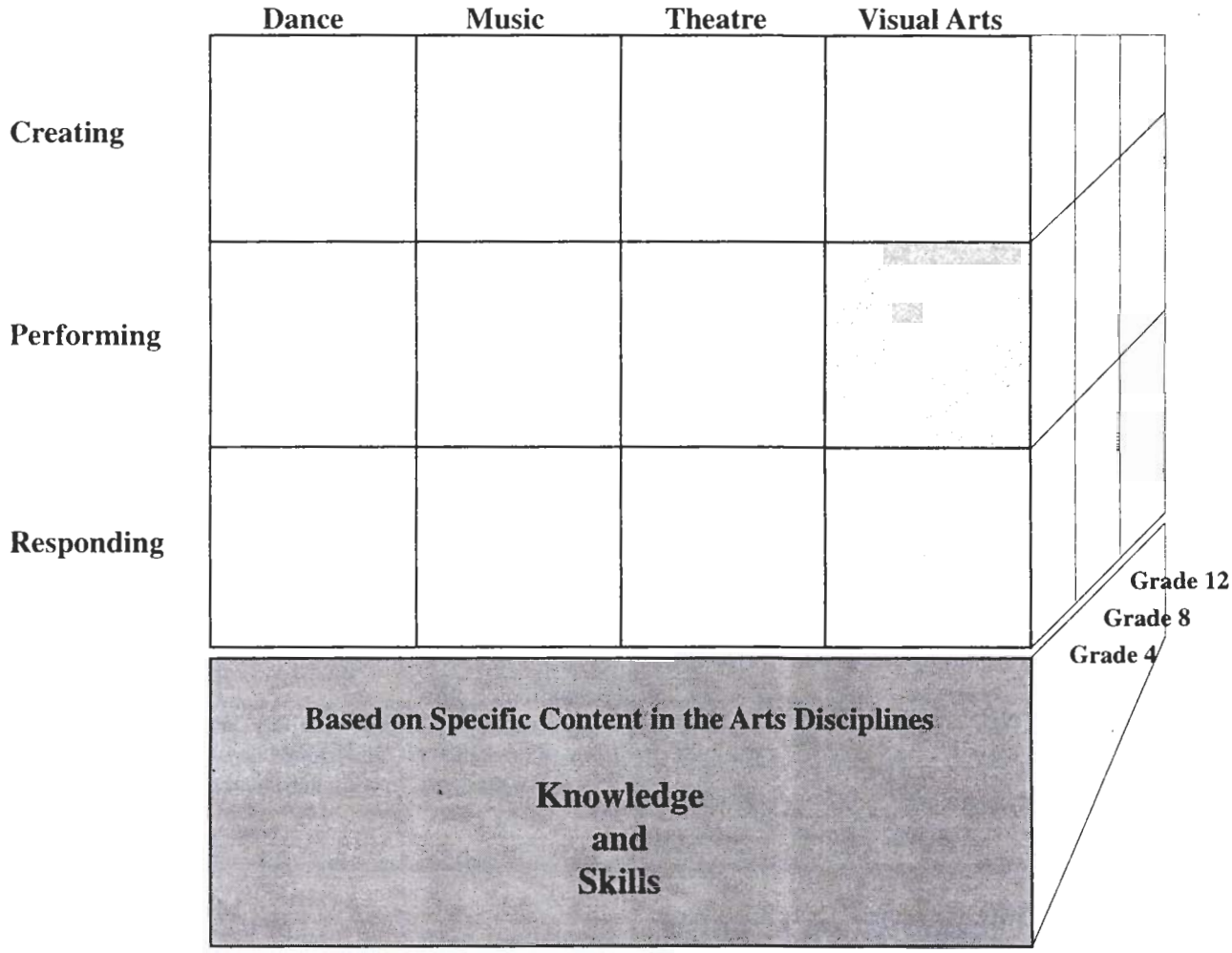
As a form of authentic assessment, teacher portfolios may play a major role in the overall evaluation of a teacher. Numerous universities, such as the University of Colorado at Boulder, Marquette University and Murray State University, now use portfolios to make personnel decisions. Many other states and institutions use teacher portfolios to augment more traditional assessment measures, such as standardized tests and observation checklists.

However, the use of teacher portfolios for high-stakes decisions, such as certification and advancement, is not universally endorsed. The reasons for caution often cited include the subjectivity involved in evaluating portfolios, the

(continued p.19 bottom)

Table 5

The Framework Matrix



(continued from p.18)

variability in content and construction of portfolios, and the lack of consensus in what a teacher should know and be able to do.

The majority of the programs that use teacher portfolio are teacher education programs. These programs use portfolios to increase reflection and provide an ongoing record of a teacher's growth. The portfolio provides a vehicle for assessing the relationship between teacher choices or actions and their outcomes. In addition, teachers are encouraged to share their portfolios, during construction, with both beginning and experienced teachers. This continuous dialogue is designed to provide a rich context in which to experience the multifaceted nature of teaching.

How Is a Teacher Portfolio Evaluated?

Portfolios that are used to make personnel decisions tend to come under a higher level of scrutiny than if the intended use is professional growth. This scrutiny is due to the importance of the consequences involved in using portfolios for personnel decisions, and has resulted in several concerns. Most often cited areas of concern are the flexibility and subjectivity of the portfolio.

The construction of a portfolio is such that each portfolio is unique and tailored to the individual. As a tool for professional development, this is a positive feature; as a tool for arriving at personnel decisions, where comparability between teachers (often from different subject areas) is desired, the lack of standardization is a problem.

The lack of, or need for, standardization can be rectified by requiring certain items in the portfolio of a teacher seeking a position or to advance. Other items may be included at the teacher's discretion. Mandated items typically include:

- Statement of teaching responsibilities.
- Statement of teaching philosophies and methodologies.
- Description of efforts to improve one's teaching.
- Representative course syllabi.

• Summary of institutional instructor evaluations by students.

The second concern of portfolio assessment, the subjectivity in the evaluation of the portfolio, is somewhat problematic. Teacher evaluation, in any form, is subjective. The question then becomes how to make the evaluation of portfolios as reliable and valid as possible, given their subjective nature.

Often, the solution is to use a Likert-type evaluation form, of predetermined qualities, based on the mandated items.

Figure 1

Sample Questions for Likert Evaluation

____ Are teaching materials relevant?

I.....I.....I.....I.....I
1 2 3 4 5

____ Does the teacher return student materials, such as tests and assignments, promptly?

I.....I.....I.....I.....I
1 2 3 4 5

____ Is the teacher actively engaged in improving his/her teaching skills?

I.....I.....I.....I.....I
1 2 3 4 5

Questions are then grouped into categories, such as Instructional Design, Course Management and Content Expertise, and weighted. Ratings may then be combined to generate categorical and/or overall ratings.

Steps for Implementing a Portfolio Program

Start slowly. Instituting portfolio assessment, either for advancement or growth, takes time. Allow one to two years for development, implementation and regulation of a portfolio program.

Gain acceptance. It is extremely important that both administrators and teachers accept the use of portfolios. If administrators do not relate the importance and usefulness of portfolios to their teachers, the project will fail. Likewise, if teachers do not value the portfolio approach, then they will not put forth the effort needed to ensure success.

Instill ownership. Teachers must be involved, from the beginning, in developing the portfolio program. They must feel ownership over the program's direction and use.

Communicate implementation. The teachers need to know, explicitly, how the portfolios will be used. If they will be used for advancement, then the expected structure and intended scoring methods need to be explained in detail.

Use models. Models of portfolios used by other institutions are readily available (see Seldin and Associates, 1993). These models may easily be adapted and provide examples for teachers developing their portfolio.

Be selective. Portfolio contains carefully selected items that reflect and substantiate a teacher's expertise and achievements.

Be realistic. Portfolios are only one form of authentic assessment. As such, they should be used as a part of the assessment process, in conjunction with other measures.

New England Conservatory of Music
290 Huntington Avenue
Boston, MA 02115
(617) 262-1120 ex. 436
(617) 262-0500 FAX

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

Address Correction Requested

Attention: Richard Colwell and John Holly, *editors*