

# Measuring Student Achievement in the Future Based on Lessons from the Past: The NAEP Arts Assessment

By Christina Schneider

In 1997, the National Assessment of Educational Progress (NAEP), also known as the Nation's Report Card, assessed eighth-grade music students throughout the United States on their ability to perform, create, and respond to music. The next assessment in music is scheduled for 2008. If this assessment is to provide a realistic and useful look at what students in the US know and can do in music, music teachers need to get involved in the process now, as the NAEP is making plans. We need to look carefully at the 1997 NAEP assessments, determine what could make the next assessment more valuable to the music education community, and make our voices heard to those who can make decisions regarding the upcoming assessment. This article provides information on the design of the NAEP and offers ideas for improving the assessment to make it more useful to the music education community.

## No Child Left Behind and the NAEP

Since the passage of the No Child Left Behind Act (NCLB) of 2001,<sup>1</sup> states are required to participate biannually in NAEP assessments in reading and mathematics for grade four and grade eight. In addition, a recent report commissioned by the National Assessment Governing Board (NAGB), the panel that oversees the NAEP, recommended that grade twelve also be included in these biannual assessments. If music is not required, how does this information affect music educators?

Although NCLB mandates biannual assessments in reading and math, it also permits voluntary NAEP assessments in other areas, including the arts, *as funds permit*. In 1997, NAEP assessed eighth-grade students nationally in music, visual arts, and theatre. (Dance wasn't assessed due to the scarcity of dance programs in public schools.) In a 1998 *Teaching Music* article about the NAEP, Paul Lehman wrote, "the arts are included in NAEP because they belong among the basic disciplines of the curriculum."<sup>2</sup>

Music educators have a role to play in ensuring that the information gathered in the 2008 assessment is useful to the music education community. To be effective advocates for successful administration of the 2008 NAEP, we need to understand the design of the test, what it is intended to assess, and the technical difficulties that prevented certain potentially useful information from being provided by the 1997 assessment.

## The Purpose and Design of the NAEP

NAEP music assessments don't provide individual student, school, or state scores. Rather, the purpose of the assessment is to provide a broad national overview of student achievement. For example, the NAEP has traditionally provided estimates of the differences in music achievement among Asian, African American, Native American, Hispanic, and white students and among students from varied socioeconomic backgrounds. Background questionnaires administered through the NAEP allow for investigation of how music instruction, including private lessons, is related to music achievement.

NAEP assessments are generally six to seven hours long. Such a long test is inappropriate to give to any one student; therefore, the NAEP divides the test into small portions called blocks. A typical test administration booklet contains two five-minute background questionnaires, one three-minute motivation questionnaire, and two twenty-five-minute blocks of content material.<sup>3</sup> Each student can complete the test

*Music educators need to speak out now if they are to have an influence on the 2008 NAEP Arts Assessment.*

administration booklet in approximately sixty-three minutes. Students in the same classroom actually take different portions of the test. This allows a long test to be administered in a short amount of time, making schools more likely to agree to participate in voluntary assessments like music.

### **The NAEP Arts Education Assessment Framework**

The NAEP's *Arts Education Assessment Framework* was developed in 1994 in conjunction with the National Standards for Arts Education.<sup>4</sup> In fact, several members of the Standards panel were also on the Framework panel. The Framework panel determined that the assessment time should be equally divided among measuring students' abilities to respond to (i.e., listening, moving, analyzing, and critiquing), create (i.e., improvising and composing), and perform (i.e., playing, singing, and conducting) music.

In our current NCLB environment, the 2008 NAEP arts assessment may not be the helpful tool it could be if music educators (as well as other arts communities and organizations) do not suggest changes in its administration.

The music education profession needs to advocate for investigation and discourse in four areas:

- researching whether students'



*Christina Schneider is an education associate in the Office of Assessment at the South Carolina Department of Education in Columbia. She can be reached at [cschneid@sde.state.sc.us](mailto:cschneid@sde.state.sc.us).*

If administered properly, the NAEP can give music educators a good idea of how well students in the US can play music.

abilities to respond to, create, and perform music are related or independent types of music achievement

- determining what data is most important and ensuring that the scores that are reported provide this data

- suggesting a policy of whom to assess in 2008

- finding out whether additional funding is needed and developing strategies for gathering those funds

### Researching Students' Abilities to Respond to, Create, and Perform Music

Research on whether responding to, creating, and performing music are related areas of music achievement can help us determine whether those three areas should be measured together on a single assessment. If the three areas are highly related (meaning that a student with a high ability to respond to music has a similar high ability to create and perform music), then measuring student abilities to respond, create, and perform on a single assessment is practical and feasible. This is because modern large assessments, such as the NAEP, use a statistical theory that's based on the assumption that a student's performance on any individual test question can be explained by his or her innate ability in the area of interest.<sup>5</sup> For scores to be meaningful, achievement tests must measure one broad type of ability. If music achievement is one holistic area of ability, then it's possible to combine the scores for each of the three areas into one composite score.

If, however, the three areas are independent types of music achievement that are not highly related to one another, it may not be practical to combine those separate types of music achievement into a composite score. The NAEP administers and reports scores on a reading test separately from scores on a writing test because students who read with a high degree of comprehension do not necessarily write with the same degree of ability. Reporting those two areas as part of the same ability would produce a test score that doesn't provide an accurate measure

of a student's ability to read or to write.

Scale scores are typically used to report student achievement in statewide and national assessments. This type of score allows policymakers and the public to compare students in different regions to one another, track trends in achievement, and determine what proportion of students are meeting state or national proficiency levels.

1997 assessment, where subscale scores were not provided for creating and performing music. By discovering why these subscale scores were not provided, we will be better able to suggest solutions to the problems that prevented such data from being available.

### Obtaining Useful Data

The 1997 arts assessment contained fourteen items measuring stu-

*Composite scale scores would provide an overall measure of music achievement.*

If the NAEP could combine responding to, creating, and performing music into a composite scale score, as was conceived in the NAEP *Arts Education Assessment Framework*,<sup>6</sup> the music education community would be able to set standards for music achievement using the NAEP terminology: basic, proficient, and advanced. The music education community would also be able to determine what proportion of students in the nation is proficient in music. In addition, composite scale scores would provide an overall measure of music achievement that would allow for comparisons of students who participate in different ensembles or who are from different ethnic backgrounds. Such information informs instruction and provides an indicator of the degree to which standards are implemented nationally.

Composite scores were not obtained in the 1997 assessment. If such information is of interest, the music education community must investigate the relationship among the three NAEP music achievement areas to determine whether a composite score would be meaningful.

NAEP typically weights subscale scores (i.e., scores in the specific areas of ability being assessed) to derive the composite score. To derive a composite music score in the 2008 assessment, we must look back at the

dents' abilities to either create or perform music and approximately fifty-six items measuring students' abilities to respond to music.<sup>7</sup> The small number of items related to creating and performing prevented the NAEP from calculating subscale scores in those two areas or a composite score combining the three areas. However, because a sufficient number of items was used to measure responding to music, the 1997 assessment could provide a subscale score for that area.

*The Role of Music Notation and Vocabulary.* The responding to music items in the 1997 assessment had to be divided into two subscales because some items depended on knowledge of music notation and vocabulary.<sup>8</sup> Those items were separated from items that didn't require music notation and vocabulary. The clustering could have occurred for two reasons. First, when music notation or music vocabulary is present in a question, the assessment may be tapping into a fourth area of music achievement, knowledge of music notation and vocabulary. (The music education community may wish to consider defining four areas of music achievement rather than the three used in 1997, but discussion of this proposal is outside the scope of this article.) Second, when music notation or vocabulary is required for a question, only students who

have received music instruction are likely to have the necessary skills to answer the question. If this is the case, the music education community may need to determine if it's reasonable to test students who haven't studied music in an area in which they haven't had the opportunity to learn the material.

*A Lack of Data on Creating and Performing.* The combined total of fourteen items in creating and performing on the 1997 music assessment was not sufficient to consistently measure each of those skills and report student achievement using a subscale score.<sup>9</sup> Only two to three of the fourteen performance items could be randomly administered to a student in a content block because complete student performances were recorded. Not enough performances could be recorded to provide a reliable subscale score of student abilities in those two areas.

Research on similar issues in science education reveals that the NAEP music assessment will likely need to use between six to twenty-three performance items related to creating music and an additional six to twenty-three performance items on performing music to obtain reliable subscale scores for creating and performing.<sup>10</sup> The science education community believes that to authentically measure students' skills, science experiments must be part of the assessment. These types of authentic experiences (science experiments and music performances) are typically called *performance tasks*.

*The Challenge of Administering Performance Tasks.* Performance tasks require students to perform or create a response rather than select a response, as on a multiple-choice question.<sup>11</sup> These tasks typically take a lot of time to administer because they require students to create music rather than select a response about music. Tests must be administered in a particular amount of time, which limits the number of performance tasks that may be used in any one assessment.<sup>12</sup>

In the NAEP, as with similar assessments, performance tasks require large amounts of time and money. There is a need to develop the stimu-

lus material for the tasks (such as background chord progressions for students to use during an improvisation or a metronome click to indicate the appropriate tempo to use during sight-reading). The stimulus material must be reproduced in a standardized form for students. Students need to be individually recorded singing or improvising, and test developers must create a rubric to score each performance task. Finally, raters must be hired and trained to score each performance consistently. This process often makes the use of large numbers of performance tasks cost prohibitive. The NAEP, however, has a rich history of using substantial numbers of items that must be scored by raters.

The music education community rightly desires to have students authentically assessed in performing and creating. The test developers rightly desire to create a music assessment that has enough items administered within a feasible, cost-effective time frame to permit the reporting of student achievement with a composite scale score. Any workable solution to this problem must take into account the desires of both of these groups.

The 1997 NAEP music assessment revealed that recording full-length performances of students improvising, singing, and playing instruments takes too much time and severely limits the music education community's ability to determine what stu-

Independently measuring student skills requires individual performance tasks in singing or playing music, much like auditions in which students perform without other students present. In an assessment such as the NAEP, the purpose is not to choose the highest achieving musicians but to investigate the range of music achievement among students. Because the NAEP serves a different purpose from an audition, many short samples of music performances may be enough to gather sufficient information about the range of music achievement in the student population in a short time.

*Shorter Samples as a Possible Solution.* One technique that has been successful in a statewide assessment of music achievement in South Carolina is the performance of four-bar-phrase performance tasks.<sup>13</sup> This assessment strategy, which could easily be applied to the NAEP, is not unfamiliar to general music teachers, who, like NAEP test administrators, often must assess a large number of students in a short amount of time. A general music teacher might listen to individual students as they echo one rhythm pattern, improvise one rhythm pattern, sing one phrase, or perform one measure on recorder. The teacher can then immediately move on and listen to the next student, rather than listening to complete performances from each student. In the same way, conductors typically

*The music education community rightly desires to have students authentically assessed in performing and creating.*

dents know and can do. Because complete performances were required in the twenty-five-minute block, there wasn't enough time to obtain multiple pieces of evidence about student abilities. All, however, is not lost. With slight modifications, it will be possible to measure students' abilities to create music and perform music in 2008.

choose important or difficult passages for students to perform individually for assessment purposes, rather than listening to individual performances of all the music to be performed in a concert.

Test administrators in South Carolina reported that administration of two four-bar-phrase performances required four to ten minutes per student.<sup>14</sup>

For NAEP purposes, it may be reasonable to expect that students could read the directions, listen to accompaniments, and prepare responses within a two- to three-minute period per performance task. This would allow eight to ten tasks to be administered within the twenty-five-minute block in 2008 rather than the two to three tasks recorded in the same amount of time in 1997. Such tactics would allow the NAEP 2008 arts assessment to record more samples of student work within the same amount of time, which would allow the NAEP to report separate subscale scores for creating and performing music.

Once the NAEP is able to effectively measure what students know and can do in music, the next issue the music education community needs to resolve is whose music achievement to measure.

### Determining Whom to Assess

In the 1997 music assessment, the eighth-grade general population was tested; therefore, the sample was a mixture of students who had and had not received music instruction. Students who were involved in music activities tended to score at the higher end of the responding-to-music scale than students in the general population.<sup>15</sup> (As indicated previously, subscale scores for creating and performing and a composite score for all three areas could not be calculated.) Although this finding did show that music instruction affects music achievement, music educators could not determine what eighth-grade music students could do regarding the National Standards because the item-level results were based on the average eighth-grade student rather than the average eighth-grade music student.

Assessing a sample of students from the general population makes it difficult to investigate what factors are related to greater achievement among music students nationally because the sample size of music students is likely to be too small to permit stable comparisons of music subgroups. The NAEP will again assess the eighth-grade general population in 2008 unless the music education

community, working with MENC, advocates otherwise. If the NAEP assesses only those students who have formally studied music, then the sample size will be large enough to allow music educators and researchers to better understand who is enrolled in music courses nationally and the music experiences those students receive.

Music educators may need to come together to develop a consensus regarding whom to assess in the 2008 music assessment. One compromise may be to suggest sampling enough students from both the general population and the music population so that results for each group may be reported. However, sampling more students increases the cost of an already expensive assessment both in terms of materials and manpower.

### Seeking Additional Funding

In our current NCLB environment, it may be necessary to find additional monies to support the 2008 arts assessment because of the nationwide emphasis on reading and mathematics. Additional funds have

been sought and used in the past for the arts assessment. For example, the National Assessment Governing Board, the policymaking body for the NAEP, received funding from the National Endowment for the Arts in conjunction with the Getty Center for Education in the Arts to develop the *Arts Education Assessment Framework*.

MENC, working on behalf of the music education community, may need to collaborate with other arts organizations to advocate for the funding and the administration of the NAEP arts assessment. With a successful 2008 assessment, the music education community will be able to learn much about what our students know and can do and what experiences affect their achievement.

### What Music Educators Can Do

The lessons learned from the 1997 arts assessment may be put to good use when preparing for the 2008 assessment only if the music education community thinks carefully about how to improve the process. The technical difficulties that did not allow for subscale and

MENC  
MENC  
MENC

## MENC Resources

After the 1997 arts assessment, MENC published the following articles related to the NAEP. You can order back issues of MENC publications by calling 800-828-0229. Articles are also available in many library periodical databases.

- Cassidy, Jane W. "The NAEP Arts Report Card: Where Do We Go from Here?" Update: *Applications of Research in Music Education* 19, no. 1 (2000): 3-8.
- Lehman, Paul R. "Making the Most of the 1997 NAEP." *Teaching Music* 6, no. 2 (1998): 30-31, 42.
- Sims, Wendy. "Why Should Music Educators Care about NAEP?" *Teaching Music* 7, no. 4 (2000): 40-45.
- "The 1997 NAEP in Arts Education: What Every Music Educator Needs to Know." *Music Educators Journal* 85, no. 1 (1998): 35-36.

Additional information about the NAEP is available at <http://www.menc.org/naep/naep.html>.

composite scale scores on the 1997 assessment need to be resolved. Perhaps MENC could develop a task force to investigate the issues and suggest possible changes for the 2008 NAEP. The music education community should suggest that this process occur immediately so that suggestions may be delivered to the National Center for Education Statistics, the entity that administers the NAEP, in time for the suggestions to be considered as the planning occurs in 2006.

It's not too early to plan for the NAEP 2008 arts assessment. Music educators in elementary and secondary schools can collaborate with music education researchers in their state by permitting researchers to measure their students' abilities to respond to, create, and perform music so that the relationships among those areas of music achievement may be determined. If music educators within districts and states work together with their state music organizations to open up their classrooms, important research can be accomplished in a short time.

Music educators (along with their colleagues in dance, drama, and the visual arts) can develop position statements regarding the types of information they can use from the arts assessment. Determining what information the music education community needs may help the NAEP answer the question of whom to assess: music students or a mixture of music and nonmusic students. As a music education community, we can read the NAEP Arts Education Assessment Framework, available at <http://www.nagb.org>, and *The NAEP 1997 Arts Report Card: Eighth Grade Findings From the National Assessment of Educational Progress*, available at <http://nces.ed.gov/nationsreportcard>.<sup>16</sup> By educating ourselves, we will be prepared to come to the table understanding the issues from the past so that we can better assess and teach our students in the future.

## Notes

1. *No Child Left Behind Act of 2001*, Public Law No. 107-110, *US Statutes at Large* 115 (2002): 1425.
2. Paul Lehman, "Making the Most of

the 1997 NAEP," *Teaching Music* 6, no. 2 (1998): 42.

3. Shi-Chang We, *Technical Report of the NAEP 1994 Trial State Assessment Program in Reading* (Washington, DC: National Center for Educational Statistics, 1996), 29.

4. National Assessment Governing Board, *Arts Education Assessment Framework* (Washington, DC: NAGB, 1994); and Coalition of National Arts Education Associations, *National Standards for Arts Education* (Reston, VA: Music Educators National Conference, 1994).

5. Ronald K. Hambleton, H. Swaminathan, and H. Jane Rogers, *Fundamentals of Item Response Theory* (Newbury Park, CA: Sage, 1991).

6. National Assessment Governing Board, *Arts Education Assessment Framework*.

7. Hillary R. Persky, Brent A. Sandene, and Janice M. Askew, *The NAEP 1997 Arts Report Card* (Washington, DC: National Center for Educational Statistics, 1998).

8. *Ibid.*, 186.

9. *Ibid.*

10. Richard J. Shavelson, Gail P. Baxter, and Xiaphong Gao, "Sampling Variability of Performance Assessments," *Journal of Educational Measurement* 30, no. 3 (1993): 215-32; and Richard J. Shavelson, Maria A. Ruiz-Primo, and Edward W. Wiley, "Note on Sources of Sampling Variability in Science Performance Assessments," *Journal of Educational Measurement* 36, no. 1 (1999): 61-71.

11. Therese Kuhs, Robert L. Johnson, Susan A. Agruso, and Diane Monrad, *Put to the Test: Tools and Techniques for Classroom Assessment* (Portsmouth, NH: Heinemann, 2001).

12. Norman E. Gronlund, *Assessment of Student Achievement*, 7th ed. (Boston: Allyn and Bacon, 2003).

13. Ching Ching Yap, Christina Schneider, Robert L. Johnson, Dawn Mазzie, and Sameano Porchea, *Documentation for the South Carolina Arts Assessment Project Year One: Fourth-Grade Music and Visual Arts Assessments* (Columbia: South Carolina Department of Education, 2003).

14. *Ibid.*

15. Persky, Sandene, and Askew, *NAEP 1997 Arts Report Card*.

16. National Assessment Governing Board, *Arts Education Assessment Framework*; and Persky, Sandene, and Askew, *The NAEP 1997 Arts Report Card*. ■

## STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION

(Act of August 12, 1970: Section 3685 Title 39, United States Code.)

1. Title of publication: *Music Educators Journal*.
  2. Date of filing: September 21, 2005.
  3. Frequency of issue: 5 times annually: January, March, May, September, and November. Subscription price: \$10.00 with membership.
  4. Complete mailing address of known office of publication: 1806 Robert Fulton Drive, Reston, VA 20191-4348.
  5. Complete mailing address of headquarters of general business offices of the publishers: 1806 Robert Fulton Drive, Reston, VA 20191-4348.
  6. Full names and complete addresses of publisher, editor, and managing editor: Publisher—MENC—The National Association for Music Education, 1806 Robert Fulton Drive, Reston, VA 20191-4348. Christine Stinson (Editor) and Teresa Preston (Managing Editor)—MENC, 1806 Robert Fulton Drive, Reston, VA 20191-4348.
  7. Owner: MENC—The National Association for Music Education. National Executive Board: (David Circle, KS; Lynn Brinckmeyer, TX; Willie L. Hill, MA; Earl Groner, NY; Janer R. Barrett, IL; Jim Rice, WA; Fran M. Page, NC; Ron Chronister, KS; George M. DeGraffenreid, CA).
  8. Known bondholders, mortgagees, and other security holders owning or holding one percent or more of total amount of bonds, mortgages, or other securities: none.
  9. For completion by nonprofit organizations authorized to mail at special rates: The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding twelve months.
  10. Extent and nature of circulation:
    - A. Total number of copies printed (net press run). Average number of copies each issue during preceding twelve months: 78,285; actual number of copies of single issue published nearest to filing date: 73,110.
    - B. Paid circulation. (1) Mail subscriptions (outside-county). Average number of copies of each issue during preceding twelve months: 74,360; actual number of copies of single issue published nearest to filing date: 70,290. (2) Mail subscriptions (in-county). Average number of copies of each issue during preceding twelve months: 0; actual number of copies of single issue published nearest to filing date: 0. (3) Sales through dealers and carriers, street vendors, counter sales, and other non-USPS distribution. Average number of copies of each issue during preceding twelve months: 525; actual number of copies of single issue published nearest to filing date: 476. (4) Other classes mailed through USPS: Average number of copies of each issue during preceding twelve months: 0; actual number of copies of single issue published nearest to filing date: 0.
    - C. Total paid circulation. Average number of copies of each issue during preceding twelve months: 74,885; actual number of copies of single issue published nearest to filing date: 70,766.
    - D. Free distribution by mail: Samples, complimentary, and other free copies. (1) Outside-county: Average number of copies of each issue during preceding twelve months: 119; actual number of copies of single issue published nearest to filing date: 123. (2) In-county: Average number of copies of each issue during preceding twelve months: 0; actual number of copies of single issue published nearest to filing date: 0. (3) Other classes mailed through USPS: Average number of copies of each issue during the preceding twelve months: 0; actual number of copies of single issue published nearest to filing date: 0.
    - E. Free distribution outside the mail (carriers or other means): Average number of copies of each issue during preceding twelve months: 200; actual number of copies of single issue published nearest to filing date: 0.
    - F. Total free distribution (sum of D and E): Average number of copies of each issue during preceding twelve months: 319; actual number of copies of single issue published nearest to filing date: 123.
    - G. Total distribution (sum of C and F): Average number of copies of each issue during preceding twelve months: 75,204; actual number of copies of single issue published nearest to filing date: 70,889.
    - H. Copies not distributed. Office use, left over, unaccounted, spoiled after printing. Average number of copies of each issue during preceding twelve months: 3081; actual number of copies of single issue published nearest to filing date: 2221.
    - I. Total (sum of G and H)—should equal net press run shown in A). Average number of copies of each issue during preceding twelve months: 78,285; actual number of copies of single issue published nearest to filing date: 73,110.
  - Percent Paid and/or Requested Circulation. (1) Average percentage of copies of each issue during preceding twelve months: 99.6%. (2) Actual percentage of copies of single issue published nearest to filing date: 99.8%.
- I certify that the statements made by me above are correct and complete.—Christine Stinson, Editor