

Assessment SRIG Newsletter

A Special Research Interest Group of MENC's Society for
Research in Music Education

No. 24
September 28, 2007

From the Editor:

Welcome once again to the Assessment SRIG newsletter. In this issue there are several items that have inspired some thought on my part, and believe they will be of interest to the SRIG membership. Here in Washington State, and on my campus in particular, issues of assessment are taking center stage. We are grappling with performance evaluation in the K-12 schools, revising music certification requirements for teachers, developing a new licensure test that will align with state learning goals, and conducting regional and national program accreditation. All this, of course, while maintaining current teaching and musical activity at a high level. I believe it is more important than ever that our conversations at the national level be frequent, broad-ranging, and occasionally visionary. If we can avoid blind alleys and the mistakes of the past, we'll be more likely to benefit students, teachers, schools, and communities.

I try to keep up with reading as best I can, but would welcome messages from the membership with hot tips on interesting items. I found this article by Debra Viadero discussing the mixed responses to the No Child Left Behind act to be worthwhile:

<http://educationweek.org/ew/articles/2007/06/20/42rand.h26.html>

All the best,

Bret Smith
Chair, Assessment SRIG

In This Issue:

- **National Collaborative on the Assessment of Student Standards**
- **Guest Commentary: Richard J. Colwell (Part Two)**
- **Contact the Editor**

1. National Collaborative on the Assessment of Student Standards

Focus of National Collaborative is on the Assessment of Student Standards

The Council of Chief State School Officers (CCSSO) is a nonpartisan, nationwide, nonprofit organization of public officials who head departments of elementary and secondary education in the United States, the District of Columbia, the Department of Defense Education Activity, and five U.S. extra-state jurisdictions. CCSSO provides leadership, advocacy, and technical assistance on major educational issues. One of the main projects of CCSSO are the SCASS member organizations. SCASS stands for: State Collaborative on Assessment and Student Standards. SCASS membership groups are focused around a topic or subject area (such as large scale assessment, assessing students with special needs, the arts, formative assessment, or social studies). States pay membership fees to participate in on-going professional development around the particular topic/subject area.

SCASS-Arts is the only state-based, nationally focused group addressing the development and refinement of arts education assessment materials for large-scale, district-level, and classroom-based assessment and professional development connected to the

National Standards in Arts Education. The SCASS-Arts Education Consortium is a collaborative venture among member states to develop and disseminate model local and state assessments in the visual and performing arts and provides professional development materials and opportunities for teachers in developing and using arts assessments.

Among other projects, the group has developed and implemented a web-based item review process. This process includes professional development for teachers in item writing leading to the submission of items to a password protected website where they are screened for content and assessment accuracy by a panel of experts. Items are judged according to criteria developed by the group, and either sent back to the originator or advanced to the final pool, which is a secured pool of high-quality arts assessment items for use by member states.

In Arizona, a member state, a recent project has been initiated on statewide arts assessment. SCASS-Arts is providing Arizona arts educators with high quality professional development on how to develop test items for assessing student knowledge and performance in music and the other art forms. Arizona hopes to use this training as the beginning step in the creation of a statewide assessment in music. SCASS-Arts will continue to assist in this process, and the national web-based item development process will be used to “vet” items created in Arizona to insure creation of high quality items for use in a statewide test.

SCASS-Arts and its member states are interested in promoting the use of quality arts assessment tools at the state, district, school, and classroom level. SCASS-Arts has multiple projects currently underway, including multi-state trainings in New England; assessment creation in Los Angeles Unified School District; development of high stakes tests in dance and theatre in New Jersey; and the potential for statewide assessment in all four art forms (dance, music, theatre and visual arts) in Arizona. SCASS-Arts welcomes questions from the field, and would like to serve as a resource for arts assessment projects as well as for arts educators interested in assessment.

For more information contact Frank Philip, CCSSO and SCASS Arts project manager at 202-336-7046 or frankp@ccsso.org or visit www.ccsso.org.

Contributed by Dr. Suzanne Burton, University of Delaware, on behalf of SCASS-Arts Steering Committee.

2. NASM Executive Director Comments

One of the Assessment SRIG sessions in Milwaukee next spring will focus on issues of accreditation of college music and music education programs. With permission, I’m reprinting the text of a policy letter by Samuel Hope, Executive Director of the National Association of Schools of Music, regarding recent attitudes toward assessment systems and program accreditation. I think all of his comments are interesting and can be found at: [http://nasm.arts-accredit.org/index.jsp?page=BRIEFING LETTERS: Spring 2007](http://nasm.arts-accredit.org/index.jsp?page=BRIEFING%20LETTERS%20Spring%202007) --Editor

NATIONAL ASSOCIATION OF SCHOOLS OF MUSIC

11250 Roger Bacon Drive, Suite 21
Reston, VA 20190
Telephone: (703) 437-0700 Facsimile: (703) 437-6312

March 21, 2007

Dear Colleagues:

The basic message of this letter is that working for results or “outcomes” in terms of disciplinary or professional content is different than using the concept of outcomes as the basis for political action to change the focus of higher education from content to quantitative

assessment, rankings, comparisons, and the images they produce, or to promote centralization. Please read on to find out why, and why understanding these issues is important for you and your institution.

In our fifth briefing letter, we talked about the relationship between ends and means. I promised to provide some explanation of why assessments, once universally thought of as means, are now regularly presented and acted upon in policy contexts as though they were ends. How have assessment and assessment results become so deeply conflated with education and educational results? What are some of the forces producing these conditions?

There are many answers. Let us start with a positive one. First, educational assessment is a field of inquiry and practice. Like other such fields, self-motivated practitioners continue to learn new things and develop new techniques. They work to make their field as influential as possible. A second factor is technology and the expectations it raises. Computers are especially good at large scale math problems and data storage and retrieval. Rapid technological advances also produce the illusion that everything could or should work just the way technology does. If we could just digitize everything, write the right program, and put it into universal circulation, our problems would be over. A third factor is managerial responsibility. The further an administrator is from the location of actual work in a large and complex delivery system, the greater the desire for accountability mechanisms that have instant credibility, especially to others. A fourth factor is the residual power of scientism in our national consciousness, a positive in the sense that it drives discovery and application in the scientific domain. We love to consider ourselves scientific, so there is a tendency to believe that anything can be explained and ordered by the empirical methods of science and manifested in a technique or technology. But pure scientism can be a negative. It often leads to a non-negotiable position: anything that cannot be quantitatively ordered and explained by the scientific mode of thought is suspect or fraudulent. This view ignores many other ways of working in and apprehending the world – artistic, historical, and intuitional for example.

Each of these four conditions and perspectives can be factored into larger more holistic efforts to improve teaching and learning and to address other goals of higher education. Assessment techniques, technologies, management assurance necessities, and scientific methods and expectations can all be used productively in judgment-based systems applied to dynamic situations, and especially where answers are local and specific rather than universal. But using each of these four conditions and perspectives to discredit such judgment-based systems produces a serious problem. Here, disciplinary content considerations are shunted aside, parts or aspects are asserted to be the whole, and local expertise is discredited.

Of course, there are far more factors involved in producing the present policy environment in American higher education. The term “outcomes” is not just a signifier of results or of using assessment techniques to serve larger educational purposes. “Outcomes” also refers to a movement with political ambitions. We prefer to talk about this movement in terms of ideas rather than people. How can the presence of this political effort be identified and separated from considerations of outcomes as results and assessment as service to content? There is a simple, reductionist answer: when the term “outcomes” is being used to create mistrust, it is usually being used in the service of political action. There are other indicators. A large set of assertions and assumptions appear regularly in the promotional arguments of the outcomes movement. The first is that educational institutions are 100% responsible for student learning. This assumption is based on a belief in a degree of sameness among individuals that cannot be sustained by empirical evidence. It also represents a denial of the mutual responsibility between institution and student; institutions are considered factories processing identical people, and producing identical products that can be monitored and compared through quantitative assessment. The outcomes movement asserts that the math problem is simple, and the simple answer, fully representational. But consider this: without the 100% formula, the mathematical aspect of the assessment analysis becomes inordinately difficult. For example, in a small institution of 2,000 students, what happens to our math problem if students are 50% responsible for their own learning? Thousands of factors must then come into the equation. The real differences among individuals in terms of talents, orientations to work, study habits and so forth must be considered along with specific institutional and program missions, and mode of thought and disciplinary differences. Some of

the factors in the equation are changing constantly. To deny reality by pretending that institutions are 100% responsible is to build a whole system of thought and action on a great fallacy and its orbiting illusions.

Here are several other themes regularly sounded by the outcomes movement. Experts in fields and disciplines cannot be trusted to evaluate because they represent special interests – their fields and the profession of teaching. All significant educational results can be determined the minute a particular program is over. High quality work is driven by coercion or fear of public embarrassment. No one cared about student learning until the advent of the outcomes movement, and to this day, no one but the outcomes movement cares sufficiently about it. The outcomes movement is the arbiter of success or failure across all of higher education, including work in disciplinary and professional content, even though outcomes movement proponents do not and could not create “outcomes” as results in more than one field, usually assessment itself.

How has so much that is so wrong, insupportable, and even nonsensical, become so accepted? The simple answer is repetition. Without a serious, visible opposition, repetitions of these notions have been doing their work to produce automatic acceptance for about 25 years. But there are many other reasons why the outcomes movement has gained so much influence. Here are a few of the many political and psychological forces at work. The outcomes movement promises simple “scientific” answers to complex questions. It promises a kind of democratic leveling through standardized evaluation and comparison. It promises power without reference to content and therefore is attractive to those who value process over content, and images over substance. It supports bureaucratic expansion and central control. It treats institutions as though they were competing factories and is thus consistent with the kind of competition that creates clear winners and losers, again a model embraced by many in our society. It meshes beautifully with the rhetoric of transparency and accountability. It is consistent with vocational goals for education, this in contrast to intellectual, artistic, and professional goals. It claims to address and reduce the escalating costs of higher education for students and their families. It is consistent with the concept of education as a business and the student as a consumer. There are many others, but one of the most powerful is that the outcomes movement both uses and is nourished by the culture of accusation, argument, and denunciation that dominates so much of our journalism and public life.

All of these issues and conditions put American higher education in a challenging place. There is reason to be concerned that this place is not a good one from which to make wise decisions that will protect and advance the full range of American higher education over the long term. Effective policy cannot be built on false assumptions, even if they are congruent with aspirations, notions, and images that seem attractive. But critiquing the outcomes movement is difficult because it has created a sound bite word prison. Any criticism is answered with the assertion that the criticizer does not want to be accountable and thus cannot be trusted. This rhetorical protection has been effective for a long time. It is used regularly against content-based professionals who object to the one-process-fits-all approach of many assessment regimes, and it strikes fear among those in higher education who must explain what the academy does to those on the outside.

The resultant failure to debate has created a significant problem. Accepting the tenets and assertions of the outcomes movement results in a significant loss of perspective. The assessment part is substituted for the educational whole. Procedure is substituted for content and reductionist indicators are substituted for educational quality. Bureaucratization and standardization of outcomes produces an anti-innovation climate. The outcomes movement constantly increases time taxes on productivity by conflating the reporting of results with the production of results. But most tragic of all, the outcomes movement produces its influence by fomenting mistrust in all educational systems and educational professionals. If outcomes assessment procedures or outcomes stewardship were ever deemed adequate, the movement would lose its reason for existence. Therefore, it cannot ever agree that outcomes efforts or reporting are sufficient. To survive, it must continually escalate its criticism and its demands.

To the extent the foregoing analysis is correct, it is easy to see why from its

beginnings, the outcomes movement attacked the accreditation system. Accreditation, traditionally conceived, is deeply concerned about student learning but works on it from a far richer and more realistic set of assumptions and practices. Specialized accreditation organizations create assessment approaches based on the natures of various disciplinary and professional contents. Engineering accreditation and arts accreditation assess differently, for example. Accreditation is able to enfold and use the concepts of outcomes as results and assessment as a service, and in fact, it had done so long before there was an outcomes movement. In response to political and public relations pressures created by the outcomes movement, many accrediting agencies have embraced outcomes rhetoric and developed programs consistent with seeking outcomes as results and using assessment techniques more effectively. But before and during the work of the U.S. Secretary of Education's Commission on the Future of Higher Education, it became apparent that all this effort was not sufficient to gain the respect much less the approbation of the outcomes movement that is operating as a political force. True to the requisites of its existence, it found higher education and accreditation guilty and escalated demands for more attention to outcomes. The permanent indictment against higher education continued.

It remains to be seen whether a sufficient number of leaders in higher education and accreditation will see the portents of this situation and create a countering force. The danger is great. Many are concerned that the Department of Education now wishes to use the federal relationship with accreditation as a means for asserting comprehensive control over higher education. But the philosophical enabler, operational driver, and public relations basis of this federal initiative is the outcomes movement and the values and assertions we have just described. Twenty-five years of promotion have come to fruition in an oblique but potentially devastating attack on educational freedom and independence. The federal connection with accreditation is just the most available means at the moment. For all those years, many in higher education, when they said "outcomes," thought they were promoting results or assessment in service of content, or using the language of the moment as a matter of rhetorical convenience. But tragically, and in most cases inadvertently, they were also giving credibility and delegating influence to a movement that has obfuscated the content-based center and achievement of higher education and sold legislators, bureaucrats, and many in business on the notions that accountability is a simple matter that can and should be quantified, standardized, and centralized, and that individual and local decisions are best replaced by outcomes specialists who are agents of central governmental control.

The next briefing letter will deal with accreditation and the public interest.

Thank you for your attention and best wishes.

Samuel Hope
NASM Executive Director

3. Guest Commentary: Richard J. Colwell (Part Two)

Richard J. Colwell is Professor Emeritus of Music Education at the University of Illinois and the New England Conservatory of Music. He is founding editor of the Bulletin of the Council for Research in Music Education, the Quarterly Journal of Music Teaching and Learning, and the newsletter of the Measurement and Evaluation SRIG. In our last issue, Dr. Colwell related some aspects of the history of the SRIG and offered several suggested topics for discussion by the membership.

I've rambled too long about assessment issues that should be discussed by music educators and I have pages more of notes. My priority concern today is the construction of tests at the local, state, and national level. Much of the effort is modeled on the 1997 NAEP. It should be obvious that the states do not have the resources to construct valid assessments. At the national level, resources have been so scarce as to affect the quality of the instrument. In

the second NAEP, assessing performance was dropped because of a lack of resources. I have the feeling that there are individuals who do not know the priority of music education if it is not the ability to perform. The third assessment also lacked resources and ETS, Pearson Educational Measurement, National Computer Systems, and Westat attempted to use various statistical procedures to estimate what arts students knew and could do. According to Mark Musick former president of the Southern Regional Education Board, NAEP standards were designed to establish what students ought to know to do well in the next grade and beyond. The first national assessment was designed to determine what students at four different age levels knew and could do. It was not related to any curriculum. A careful reading of the technical manual of the 1997 assessment indicates from the design that this assessment only determined what 8th grade students knew and could do. The consultants obviously wanted the assessment to be closely related to a curriculum based on the voluntary national standards. This limited the information in the 1997 assessment to one grade (as compared to the four ages in the first assessment) without accomplishing the objective of the consultants. State standards more typically reflect what teachers say are the levels that good students reach in their classes. The purpose is important and we can learn from the struggles in math, social studies, and language arts. We can also learn about standards and revising standards from math, language arts, social studies, and now science. These subjects have revised their standards at least three times and one would hope in relationship to the changing purposes of NAEP and/or the three different formats (new, state, and trend) of NAEP.

Whether or not the voluntary national standards should be the basis of the music NAEP is a reasonable point for discussion. David Elliott (2006) states that the present standards are not related to any critically reasoned statement of musical values and are not informed by serious research (Music education and assessment: Issues and suggestions in Philip Taylor [Ed.], *Assessment in arts education*, Portsmouth: Heinemann).

Sam Hope (2006) raises the question in a different way when he describes the characteristics of a "basic" subject. Based upon his 16 characteristics, there is a serious question about the present justification that music is a basic or core subject. (Characteristics of a "basic" discipline in elementary and secondary education, *Arts Education Policy Review*, 108 [2], 3-5). Researchers attempting to justify music instruction such as James Catterall are unconcerned about the content of music instruction, band and general music are alike, looking only for effect in any experience.

The present state and national standards are neither an adequate basis for a curriculum or a state or national test. Their contribution according to Cronbach lies mostly in the symbolism of their existence. And Brent Wilson, a visual arts educator, argues that individual standards do not form a coherent vision of the purposes of arts education. Wilson (1992) has been active in visual arts assessment informing his clientele that assessment has multiple roles: criticism, grading, qualification, placement, prediction, diagnostic, didactic feedback, communication, accountability, representation, innovation, implementation, and curriculum maintenance (*Design for Arts in Education*, 93 [3], 34-44). Some standards are very broad and others narrow and no rationale is given for this. Many are ambiguous, there is lack of a routine mechanism to clarify the limits of the standards, and a limited tradition of vigorous and ongoing debates about their test use. The standards are essentially moral prescriptions and are easier to state than to follow. Some, but far from all state standards specify grade-by-grade content in core subjects. In some states, content standards are little more than category labels describing collections of curricular aims in particular content areas. Heidi Glidden and Amy Hightower (2007) looked at 714 standards and found a match with assessment on about 70%. In one state 81% of tenth-grade standards were repeated: In reading 42% were repeated from grade 2. Their other examples make for interesting reading. (*American Education*, 31 (1) 24-33; 50-51.) State and national standards are judgments carried out by reasonable people in a social and political context. Judgments embody trade-offs, not truths, and are based on education, values, and personality.

The 1997 arts assessment is an impressive document and ETS and its supporting units did a remarkable job considering the financial restraints. I cannot imagine how state officials believe they can develop comparable instruments as ETS had at least 11 million dollars in start-up funds for test development. There is a strange reliance on responding, creating, and

performing as guiding principles, almost like a taxonomy. Knowing, valuing, and a host of other principles are surely of equal or greater importance. The authors would have profited had they used one of the taxonomies to guide their thinking as scoring was often stopped when a student had not "reached" a proficiency level, a concept useable in IRT (individual response theory) but not yet feasible in music. ("A student's response for a given question was classified as not reached if a student failed to answer the question and all others following it in a given test booklet or Creating/Performing block," p. 6 of *The NAEP 1997 Arts Technical Analysis Report*, March 2004, ETS). The three principles did not work and performing and creating had to be combined in some analyses and, in music, separate analyses were required when notation was involved. IRT was attempted only with responding as the sample for Creating/Performing was inadequate. IRT works better for groups than for individuals. This is of interest as the arts assessment design was based on individuals and had schools been of interest, a different sampling design would be called for.

The sample size was 2275 students, 1999 from public schools. I'm not impressed. As a comparison, I looked at sample sizes of research studies reported in the most recent issues of the *American Journal of Education*: 112 (4), 113 (1 and 2). Carbonaro used 21,000; Lewis and Cheng 665 principals; Riegle-Crumb 20,745; Ma 3116; Kerr, 140,000; Kurlaender and Yun 15,800, with similar sample sizes for Massey, Smith, and Vergari. One has more confidence in making national generalizations when the sample size is adequate. The test was administered, when possible, in groups of 30 students in a non-music class. Students currently engaged in some type of musical activity took two Creating and/or performing blocks, one for the general student sample and one for students with special music knowledge. (This reduces the number of different students involved.)

The arts are the only NAEP that cannot be reported in terms of basic, proficient, and advanced. This inability sends a signal that the difficulty of items remains unknown and casts doubt on the entire philosophy of matrix sampling. With matrix sampling, students do not have an equal probability of being selected. This works in other subjects with large numbers but could not when ETS attempted to assemble classes of 30 eighth grade students in a single school. There were five Creating and Performing blocks and four Responding blocks, the creating and performing blocks divided into three creating and performing blocks for students in the general population and two creating and performing blocks for students currently active in music. ETS had only 21 multiple-choice questions and 35 constructed response questions to use in the matrix sample. There is no report of any item analysis to assist in dividing these few responding questions into equivalent blocks. To the best of my knowledge there has been only one published test in music that provided item analysis data and little, if any, additional research. Even with item analysis data on the multiple-choice questions, I would be reluctant to suggest that two items with a difficulty index of .50 were equivalent for assessing what students know and can do. ETS was extremely optimistic and administered 37 demographic questions in music and many more common to the arts for possible interpretation of the data. This really became impossible as the sample size for creating/performing was no larger than 64 and when divided for notation and technical vocabulary, even smaller.

Constructed response items had two, three, four, or five categories, an interesting idea but one that complicates interpreting any results. An attempt was made to use "parameter estimates from the initial solution as starting values for subsequent estimation in which the ability distribution over subjects was unconstrained and estimated concurrently with item parameter estimates" (p. 31). To do this, there was an assumption that most items were independent from one another, a risky assumption. ETS did create clustered items with this lack of independence on 103 items. 13 items had to be dropped due to lack of fit and 24 were collapsed to improve item fit. The reliability for polytomously scored items ranged from .60 to .91, considerably lower than reliabilities in Reading, Mathematics, and Writing. Pearson trained the scorers and one can imagine that the sample provided for level 1, 2, and 3 established the scoring level. Those of us experienced in music know that a performing level of 1 can differ considerably from one community to the next. ETS claims to have used item anchoring in establishing equivalency. To use item anchoring, an item should be answered correctly by 65% of the population. This obviously, did not happen. ETS also used jack-knifing for variance estimation, an accepted statistic but seemingly not applicable with the item blocks

used.

The paper-and pencil component for each session required 1½ to 2 hours. Performance activities required 20-60 minutes per student. Performance testing began shortly after the paper and pencil session and when there was a large number of students, continued into the next day. It took two or three days to complete all assessment activities in a school. One admires such thoroughness but it hardly establishes a model for local and state assessment activities and it would be a stretch to consider such testing "authentic." This Herculean effort requires a scholarly discussion among SRIG members and perhaps advice based upon our experience in music assessment to any future large-scale assessment effort.

It is problematic when students are used as the unit of analysis when a curriculum is experienced by students in a group, under a teacher, as this makes the classroom the smallest independent unit of analysis.

NAEP is said to be a reasonable standard for comparison among schools, districts, and states because its content frameworks contain academic content and skills that appear in state standards. In my opinion there would be a problem with using the 1997 NAEP or a similar test as a national benchmark due to some of the issues I have raised. We need a test with items that are comparable across schools and classrooms. Just think how varied the music experiences are for students in band, orchestra, choir, class guitar, or whatever during the four years of high school. (Of interest is that a probe for a 12th grade assessment in music was sent to these same schools but no information has been released on the results of this probe.)

The assumptions made by ETS for the music assessment are similar to assumptions made in international comparisons. Professor Jean-Yves Rochex (2006) tested these assumptions with an intensive case study in language arts. If there are issues with difficulty, knowing, and understanding in language arts, these same issues are multiplied in music. Rochex's article "Social, methodological and theoretical issues regarding assessment: Lessons from a secondary analysis of PISA 2002 Literacy tests" (in Judith Green and Allan Luke [Eds.], *Review of Research in Education* 30, 163-212. Available at <http://rre.sagepub.com/content/vol30/issue1/>) makes the following points.

PISA assessments are designed to assess the aptitude to undertake tasks found in everyday life, not the subjects of the school curriculum. (The 1997 music test was given to students not enrolled in music.) Can literacy be evaluated outside its cultural, linguistic, and even local context? PISA worked on the idea that literacy or numeracy is universal, independent of cultural and linguistic contexts. They also thought them to be stable and consistent for a given individual regardless of the situation and of the test in which they are measured (p. 180). An aim is to ensure that individual performance assessment is independent of the difficulty evaluation of items; IRT presupposes that whatever items are passed by an individual, one can make the same assessment of his or her competencies on other items and that whatever group of individuals pass a test, whatever the contexts in which these individuals live, and whatever the situations in which they are assessed, one could make the same evaluation of the test's difficulty (p. 180). IRT also depends on a twofold hypothesis or validity conditions. First, whatever items a person passes, he or she should obtain the same level of performance on the global scale or the three literacy subscales. Second, whatever the person's level of performance, the processes and ways of answering a question should be similar and could be organized on a continuum (p. 181).

Test scores, especially aggregate scores, possess only relative, not absolute properties or they function as hypotheses about student knowledge and competencies which are always embedded in different contexts, rather than statements of fact representing actual student competencies (p. 182). It is always necessary to consider people's intra-individuality as well as their dynamic and contextual nature. Some are motivated more by psychometric and political concerns than by theoretical reasons or research results in the different disciplines like the science of education. Case studies often have found that it is difficult to predict how pupils would answer a particular item on the basis of their response to another item (p. 185). By carefully interviewing the students who had taken the test, Rochex found that competencies

are not univocal nor stabilized contrary to the hypothesis of the PISA designers (p. 185).

How students relate to the world, to values and opinions, and to intellectual habits allows them to interpret the world and the experiences they had of the world that interfere with their reading of the texts (p. 187). As a musician, I am reasonably sure that these same factors and more affect how one understands and derives meaning from musical experiences.

As we don't have cut scores, we can ignore the issue of teachers working only with getting marginal students to the cut score level and ignoring all others.

States can't duplicate the work of ETS because of heavy time requirements, high costs of scoring, difficulty in achieving reliable scores, narrow scope of skills that can feasibly be assessed and lack of norms for comparisons. In general, performance tests are inefficient, costly and often of dubious reliability. There are good reasons for introducing a national test, including the economic and the pragmatic and states are looking to form partnerships in developing common tests but these reasons seem insufficient at present. There is no reason to not consider no or limited assessment. I know that quality assessment is a linchpin in tracking school performance and spurring improvement but it just might be that it is inappropriate for music. Our communities and our teachers have definite ideas about what a musically educated person should know and be able to do and this diversity can be a strength of music education as long as these diverse experiences are quality musical experiences.

This has been a lengthy conversation without discussing program evaluation, rating scales, peer evaluation, self-evaluation, and on and on. Professor Smith should have no difficulty in securing material that will inspire all of us to think more deeply about the potential of music assessment.

4. Contact the Editor

Editor: Bret P. Smith, Ph.D.
Assistant Professor, Music Education/Strings
Department of Music
Central Washington University
400 E. University Way
Ellensburg, WA 98926-7458

(509) 963-1548 ofc
(509) 963-1239 fax
bpsmith@cwu.edu

This work is licensed under the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.



CENTRAL WASHINGTON UNIVERSITY

Your future is Central.