

# ASSESSMENT

## Special Research Interest Group

*Fall Issue 2012*

This year has been important in regards to issues of assessment; in particular, measurement of student achievement and teacher evaluation. This issue features articles by Edward P. Asmus, William Koch, and updates from our members Jeffrey Ward, Dan Massoth, and Lynn Corbin.

### CALL FOR CONTRIBUTIONS and INVITATION TO JOIN:

- The NAFME Assessment SRIG publishes a newsletter twice yearly when contributions warrant, in Fall and in Spring. We accept articles, book reviews, and news items up to 1,000 words in length. Due dates: **March 1** for the Spring issue; **October 1** for the Fall issue. Contributions for the newsletter are welcome at any time.
- Invite your colleagues to join the Assessment Special Research Group. To join, they simply go to the NAFME SRIG website and follow the directions:  
<http://musiced.nafme.org/resources/special-research-interest-groups-srigs>
- Our NAFME SRME SRIG Liaison is Marie McCarthy, School of Music, Theatre & Dance. University of Michigan, 1100 Baits Drive, Ann Arbor, MI 48109, Ph: (734) 647-0557. She can be reached via email at [mfmcc@umich.edu](mailto:mfmcc@umich.edu).

## VOLUME 27

### Contents

RESEARCH Edward P. Asmus Assessment Ethics	2
PRACTICE William Koch A Systematic Curricular Approach to Student Assessment	5
4 <sup>th</sup> International Symposium of Assessment in Music Education	7
MEMBER UPDATES:  Jeffrey Ward Dan Massoth Lynn Corbin	8

**Website :** <http://assessmentsrig.weebly.com/>

---

## Assessment Ethics

Edward P. Asmus  
University of Miami

*The ethical foundations of test development are considered in light of established codes of ethics and standards developed by major professional organizations. In today's world where many individuals across our country are finding themselves involved in the development of statewide measures of music, a clear understanding of the ethical expectations and the established standards of test development is a must to avoid many possible pitfalls. Fortunately, these guidelines are readily available, easily understood, and greatly help the test development process.*

### Assessment Ethics

We are in a time when those in the field of music education with expertise in assessment are being asked to produce or lead the effort to produce the measurement tools that will be used to assess children's musical skills and knowledge in various states across the nation. It is not uncommon for these development efforts to be conducted outside the normal procedures that would be used in math, science, reading, and writing with usually small budgets, wide grade level and content coverage, large expectations for positive outcomes, very short time spans until full implementation, and tremendous pressure from professional organizations. These factors can lead to less than desirable results and even be harmful to the children and the teachers that these tools were designed to help.

It is the intent of this article to consider the ethical implications of the difficult situations that can exist when one accepts a test development assignment under the conditions described above. It is imperative that the developer be fully aware of the potential ethical violations that such pressure packed situations might encourage and to avoid them at all costs. There is help. There are quite a few guidelines that can help assure that the developer is following accepted procedures in the development process. In general, these can be divided into standards for test development and codes of conduct.

### Codes of Conduct

Codes of conduct that have direct bearing on the test development process have been developed by the American Educational Research Association (2011), the American Psychological Association (2010), and the National Council on Measurement in Education (1995). The purpose of these codes of conduct are summed up well in the second paragraph of the *Code of Professional Responsibilities in Educational Measurement* by the National Council on Measurement in Education (NCME).

The purpose of the Code of Professional Responsibilities in Educational Measurement, hereinafter referred to as the Code, is to guide the conduct of NCME members who are involved in any tinge of assessment activity in education. NCME is also providing this Code as a public service for all individuals who are engaged in educational assessment activities in the hope that these activities will be conducted in a professionally responsible manner. (NCME, 2010).

The *Code of Ethics* of the American Educational Research Association (AERA) is divided into two main sections: principles and ethical standards. The five principles are (a) professional competence; (b) integrity; (c) professional, scientific, and scholarly responsibility; (d) respect for people's rights, dignity, and diversity; and (e) social responsibility. There are 22 ethical standards of particular interest for test developers are four of the first five standards (1) scientific, scholarly, and professional standards, (2) competence, (3) use and misuse of expertise, and (5) avoiding harm.

The American Psychological Association (APA) ethics statement is divided into principles and standards like the AERA Code. The five general principles of the APA Code are (a) beneficence and nonmaleficence; (b) fidelity and responsibility; (c) integrity; (d) justice; and (e) respect for people's rights and dignity. There are 10 standards with 3 of direct import to the test development process: (2) competence; (8) research and publication; and (9) assessment. There are 11 different subsections to the Standard 9: Assessment of which (.03) informed consent in assessment; (.04) release of test data; (.05) test construction; (.06) interpreting assessment results; (.07) assessment by unqualified persons; (.09) test scoring and interpretation services; (.10) explaining assessment results; and (.11) maintaining test security are particularly directed at the on-going activities of the test developer.

The NCME Code has four main sections that breakdown the needs for different interests in the field of assessment: (a) general responsibilities; (b) responsibilities of those who develop assessment products and services; (c) responsibilities of those who market and sell assessment products and services; (d) responsibilities of those who select assessment products and services. This division provides useful guidance for those who develop music assessment tools because we often face the dual role of developing the tools and then having to persuade music teachers to use the tools. The NCME Code provides for this duality.

Green, Johnson, Kim, & Pope, 2007 (2007, p. 1001) provide an interesting synthesis that speaks directly to the most salient aspects of these codes. These authors established two guiding and overriding ethical principles for assessment: (1) do no harm; and (2) avoid score pollution. The former is the concept that no harm should come to anyone as a result of any facet of an assessment. No one should be hurt, embarrassed, suffer undue stress, etc. This includes that students are provided adequate preparation for being successful on an assessment. "Avoid score pollution" occurs when any activity leads to enhanced assessment performance without concurrent increase in mastery of the content being assessed. That is, the assessment scores have been "polluted." These concise and easily remembered principles are useful in avoiding a great many assessment pitfalls, but more focused guidelines that focus directly on the test development process would be useful.

### Test Development Standards

The most comprehensive test development standards available are those that were jointly developed by AERA, APA, and NCME titled *The Standards for Educational and Psychological Testing* (2010). These standards are in three parts: (1) test construction, evaluation, and documentation; (2) fairness in testing; and (3) testing applications. The first part covers validity; reliability and errors of measurement; test development and revision; scales, norms, and reporting; and supporting documentation for tests. Part two includes fairness in testing and test use; the rights and responsibilities of test takers; psychological testing and assessment; educational testing and assessment; testing in employment and credentialing; and testing in program evaluation and public policy.

The Joint Committee on Testing Practices that is comprised of AERA, APA, and NCME has developed the *Code of Fair Testing Practices in Education* (2004). This document addresses directly two separate audiences: test developers and test users. Each of the four sections of the document are divided into two parts. The first part addresses test developers and the second part addresses test users. The four sections are (a) developing and selecting appropriate tests; (b) administering and scoring tests; (c) reporting and interpreting test results; and (d) informing test takers.

### Importance of the Codes and Standards

The test developer can readily see from the comprehensive nature of just the major codes and standards cited here that the ethics of test development is considered a very important topic by the major professional organizations that focus on this subject. It is important that test developers be aware of these professional guidelines and adhere to them. In the world of high-stakes testing where the results of tests can have serious ramifications for teachers, students, and music programs in general, litigation is always a possibility.

### Reliability and Validity

Many of the ethical considerations in test development are in some sense issues of validity. According to classical test theory, reliability sets the upper bound of validity (Lord & Novick, 1968, p. 72; Salvia & Ysseldyke, 1999, p. 155). This leads us to conclude that a valid test is reliable, unreliable tests cannot be valid, and reliable tests may or may not be valid. However, without adequate reliability we have nothing. The acceptable levels of reliability in statewide testing must be higher than those often found in research studies in our field where the authors claim something along the lines of “the reliability was sufficient for research purposes.” To place a test with insufficient reliability into use, would be a great violation of ethical behavior.

### Conclusion

A word of caution is in order about our field of music education. A consideration of the two National Assessment’s of Educational Progress that included music can lead to an interesting possibility. The field thinks that it is providing students with an amazing array of skill development and knowledge. Yet, when we test for these, students don’t seem to know them. Using teachers to determine test content may lead to a violation of Green et al’s first principle “Do No Harm.” While our teachers think that something is taught, it may not be. Just how much theory does a choir student learn when the focus is on learning the music and the words for the next concert? The students, in such cases, have not been provided the opportunity to learn the material covered by the test.

The ethics of testing is one that needs to be given considerable thought by all who develop or use tests in the assessment process. Failure to follow acceptable practice standards and professional codes of ethics could lead to dramatic consequences for the test developer. It is a fairly simple task to review and follow the codes of ethics and professional practice standards major professional organizations have established. This should lead to a more relaxed environment for the test developer and better-developed tests.

### References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2010). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association. (2011). *Code of ethics*. Washington, DC: American Educational Research Association. Available at <http://www.aera.net/AboutAERA/AERARulesPolicies/CodeofEthics/tabid/10200/Default.aspx>
- American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct*. Available at <http://www.apa.org/ethics/code/index.aspx>
- Green, S. K., Johnson, R. L., Kim, D., & Pope, N. S. (2007). Ethics in classroom assessment practices: Issues and attitudes. *Teaching and Teacher Education*, 23(7), 999-1011.
- Joint Committee on Testing Practices. (2004). *Code of fair testing practices in education*. Washington, DC: Author. Available at <http://www.apa.org/science/programs/testing/fair-testing.pdf>
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- National Council on Measurement in Education. (1995). *Code of professional responsibilities in educational measurement*. Available at <http://ncme.org/resource-center/code-of-professional-responsibilities-in-educational-measurement/>
- Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (8<sup>th</sup> ed.). Boston: Houghton Mifflin.

## A Systematic Curricular Approach to Student Assessment

William Koch

Koch and Associates Educational Consulting

Music educators throughout the country are finding it difficult to meet the demands of the new teacher evaluation system. If you are looking for a curriculum that uses a systematic approach that ties what you are teaching in the classroom to student growth, read further.

The Band program at Niles West had a one-of-a-kind Instrumental Music Curriculum that was field tested for the past 15 years. Our systematic approach held each student accountable for their own continuous development of technical skills and musical knowledge. Though this curriculum was designed for a band, it can be adapted for a choir or an orchestra as well. Our curriculum came about out of desperation! From 1983-1996, our fine arts students received either a pass or fail on their report card. After years of discussions within the fine arts staff about the pros and cons of grades, we were granted permission by the school board to give “weighted” grades. Then the reality of the situation hit us—*now what?*

So, for the first year I did what many music educators do and assigned grades based on daily participation, attitude, and attendance. However, I felt there had to be a better way to assign grades. With the help of our private teaching

staff, we put together a comprehensive Instrumental Music Curriculum. The curriculum that we developed was aligned with our stated District Learning Objectives and State of Illinois Learning Standards.

Our systematic curricular approach held music students accountable for their individual performances and allowed me to track student growth. The data collection system made it easy to report the individual progress of each student while allowing me to reflect on and improve my teaching. The Instrumental Music Curriculum was comprised of a progressive etude and solo curriculum, scale study curriculum, music theory, daily demonstration of ensemble knowledge and skills, and written concert critiques.

The centerpiece of the Instrumental Music Curriculum was a series of progressive etudes and solos designed to help each student develop their musical skills and knowledge required to perform in a high school band program and beyond. Each “instrument family” had their own set of etudes and solos with each being more difficult than the one before it. During a testing week, the students would rotate out of class to be assessed by one of our para-professional staff members. Each staff member would receive inter-rater reliability training at the start of every school year to calibrate their assessment standard. The students were authentically assessed on 6 etudes, 2 solos, and 12 scale studies per year. Within a couple of days, the testing was over and the grades were entered into the electronic grade book. When the etude assessment forms were returned to the students, the teacher would change seating based on their scores.

Then the students would take out their cell phones and enter the individual category scores into a document found on our band website. Those scores were then put into a spreadsheet and sorted by student and by section where we could track the student and section category scores (tone, intonation, rhythm, etc.). You might be wondering about the legitimacy of the students entering their own category scores. Well, their “official” score for their grade was already entered into my grade book so they had no reason not to enter the correct scores from the assessment form. It was easy to corroborate the scores they entered into the database by simply referring to the spreadsheet after it was sorted by section. Besides, the students all knew that the numbers that they were entering electronically into the document on the website was to give me the ability to help make them and the band program improve.

At a quick glance, we could determine the exact performance level of each student and each section of the band. If the numbers were low for an individual student or a particular section, we could immediately address it with an intervention. That intervention might be as simple as talking to the student with a low score in tone or intonation to ascertain the problem. The student might have been playing on a bad reed that day or that the student might need to spend some time with one of our student leaders with a tuner in a practice room. If a trend was noticed in a particular section, an intervention could be created based on actual data, not just on a hunch or a “feeling.” For example, it might be decided that we need to spend more time on chorales rather than technical studies in a particular band. We used actual data that was collected at regular intervals to drive the instruction in the classroom. The new mandates on teacher evaluation are sweeping the country like a wild fire and are affecting every music teacher. If you feel frustrated and confused, you are not alone. I sincerely hope that some of the ideas presented here will help teachers generate ideas on how they can evaluate student learning and show student growth in their classrooms. Let’s keep the lines of communication open and we will all come through this together.

## 4<sup>th</sup> International Symposium on Assessment in Music Education

The music education faculties of the National Taiwan Normal University and the University of Florida School of Music will host a Symposium on Assessment in Music Education April 10-13, 2013 on the campus of the National Taiwan Normal University in Taipei, Taiwan. The purpose of the symposium is to bring together music education professionals worldwide to share the latest research, thought, and practice in music education assessment. We invite primary and secondary school music educators, higher education professionals and music education researchers, national, state and local education officials from across the world to join us in Taipei.

The Key Questions for the Symposium are:

1. **Practice.** What practices are most successful in measuring student achievement and skill in music in diverse educational systems?

2. **Measurement.** In what ways are music educators measuring student achievement in diverse musical cultures? How are these measures validated and examined for reliability?

3. **Policy.** In what ways has law and policy impacted music assessment in the world's diverse educational systems?



### Keynote Speakers:

Dr. Pamela Burnard, Cambridge University,  
United Kingdom

Dr. Richard Colwell, Professor Emeritus,  
University of Illinois, USA

Dr. Yao-Ting Sung, National Taiwan  
Normal University, Taiwan



### Chairs

Dr. Timothy S. Brophy, University of  
Florida, Gainesville, USA

Dr. Mei-Ling Lai, National Taiwan Normal  
University, Taipei, Taiwan

Dr. Ming-Hui Lin, National Taiwan,  
Normal University, Taipei, Taiwan

For more information, please go to:  
<http://conferences.dce.ufl.edu/isame/>

## Updates from Members

Dr. Jeffrey Ward is the Interim Associate Dean for Assessment and Associate Professor in the College of Fine Arts and Communication at East Carolina University.

A survey regarding outcomes assessment practices in higher education music programs was completed by 309 faculty and administrators throughout the United States. Through this survey, respondents described the following issues related to outcomes assessment:

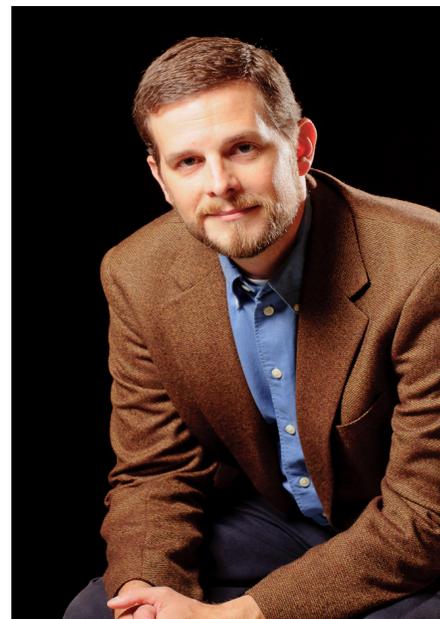
- Assessment Plan Content
- Outcomes Assessment Training
- Outcomes Assessment Benchmarks
- Outcomes Assessment Communication
- Outcomes Assessment Cycle
- Outcomes Assessment Personnel
- Barriers to Outcomes Assessment
- Resources Dedicated to Outcomes Assessment
- Outcomes Assessment-Based Decision Making

Based on the data, I have initially concluded that faculty involvement in the outcomes assessment process is necessary for an effective process because assessment is about teaching and learning (Gray, 2002). In this survey, only 17% of respondents (n=236) identified faculty as involved in the outcomes assessment process with 54% reporting involvement by an assessment chair in the music unit. If faculty feel that the assessment process is a “top-down” process, it is unlikely that significant change to teaching and learning will occur. Outcomes assessment will be viewed as a duty outside of teaching and not be incorporated into a “Cycle of Assessment Informed Instructional Design” (Ward and High, 2010, p. 506). If faculty are to be involved in the outcomes assessment process, they must be given time; yet, only 14% of respondents (n=215) reported that outcomes assessment activities is given credit to faculty toward tenure and promotion. If left to administration, the collection of assessment data will not offer the picture of student achievement necessary to improve program offerings and instruction. I am continuing

to analyze this data according to a variety of factors, including institution size, focus, accreditation, and degree programs and respondent role (i.e., faculty, administrator, assessment officer).

### References

- Gray, P. J. (2002). The roots of assessment: Tensions, solutions, and research directions. In T.W. Banta (Ed.), *Building a Scholarship of Assessment* (pp. 49-66). San Francisco, CA: Jossey-Bass.
- Ward, J. & High, L. (2010). Theory into practice: Teaching assessment strategies to pre-service teachers through a third-through-fifth-grade vocal music laboratory. In T. S. Brophy (Ed.), *The practice of assessment in music education: frameworks, models, and designs proceedings of the 2009 Florida Symposium on Assessment in Music Education* (pp. 499-512). Chicago, IL: GIA Publications, Inc.



## Updates from Members



Dan Massoth – VP of Education Solutions. MakeMusic, Inc.

### **New Standards and Assessment website resource available**

MakeMusic has developed a website containing comprehensive information about state standards as well as valuable details about how the standards are assessed. Through the collaboration with researchers, educators and administrators this website was designed to be a one-stop shop resource for up-to-date information from around the country. Additionally, materials from this summer's National Symposium on Music Assessment and Teacher Evaluation are currently available on the site and more information is being updated and added continuously. Visit <http://musicstandards.org> to learn more.

Lynn Corbin, Professor of Music, Valdosta State University  
Coordinator of Music Education

Lynn Corbin is investigating specific methods for assessing dispositions for music education students. The notion of predicting success for teachers has focused primarily on content and demonstrations of pedagogy, but the reality is that the person's personality, values, work ethic, and commitment have as much to do with success as the academic does. As the literature on successful models of assessment of dispositions is limited, Corbin decided to start asking program coordinators at various genres and sizes of teacher preparation institutions specifically how they undertook this process. A set of questions were developed and validated. Approximately 40 institutions were contacted and 21 agreed to participate. Preliminary findings are that with only one intriguing exception, all colleges and universities engaged in preparing music educators utilize some form of disposition assessment. The format, stakes, timing, and personnel participation vary widely.



*For comments, suggestions, or future submissions to this newsletter, please contact the Editor and  
Chair of the SRME Assessment SRIG*

---

**KELLY A. PARKES, EDITOR**

---

Associate Professor  
322B War Memorial Hall, Virginia Tech  
Blacksburg, VA24061-0313

Ph: 540 231 0765 Email: [kparkes@vt.edu](mailto:kparkes@vt.edu)

---