

Assessment SRIG Newsletter

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

No. 23
June 12, 2007

From the Editor:

Welcome to another issue of the Assessment SRIG newsletter. As editor, I hope to be able to produce newsletters that are interesting, thought-provoking, and timely—and as I get better at the desktop publishing aspects, perhaps a little fancier-looking. The current format seems to be taxing my abilities to the limit. I'm happy to be able to present the first half of a commentary by editor-emeritus Dr. Richard Colwell, and a feature article by Dr. Kelly Parkes.

As the editor of this newsletter stated in issue 4, April 1988, "Your contributions to the newsletter are most welcome. There are no restrictions regarding the content as long as it is of interest to the SRIG. Please send us news of research completed, sources of special reports, requests for information about or assistance with research questions, general comments and ideas or any announcements which might be of interest to members of the SRIG." Please contact me regarding style and formatting prior to submitting material.

I encourage members to respond to issues raised in the newsletter, and will create a regular "Members Respond" segment in future issues.

All the best,

Bret P. Smith
Chair, Assessment SRIG

In This Issue:

- **Assessment SRIG Website Launched**
- **SRIG Sessions at 2008 MENC National Conference**
- **Guest Commentary: Richard J. Colwell**
- **Feature Article: Kelly A. Parkes, *Assessment in the Applied Studio***
- **MENC Draft Statement on Assessment**
- **Contact the Editor**

1. Assessment SRIG Website Launched

SRIG Chair-elect Dr. Bill Bauer has been diligently working on a website for the Assessment SRIG. While we are still developing format and content, readers can visit the site at <http://assessment.webhop.org> and see what we have so far. We have located and scanned copies of past newsletters, and hope they will be of value. Please contact Bill at william.bauer@case.edu with suggestions.

2. SRIG Sessions at 2008 MENC National Conference

The SRIG will host two sessions at the MENC National Conference in Milwaukee, WI April 9-12, 2008. The first will be a brief business meeting with the primary task that of designating a new chair-elect to serve from 2010-2012. I will try to conduct this through an electronic poll of members to permit discussion of other issues. Dr. Tim Brophy of the University of Florida has graciously offered to present a summary of the symposium discussion

groups from the recent conference *Integrating Curriculum, Theory and Practice: A Symposium on Assessment in Music Education* held in Gainesville (<http://conferences.dce.ufl.edu/isame/>).

The second session will be a panel consisting of Drs. Bill Bauer (Case Western Reserve University), Suzanne Burton (University of Delaware), and Peter McAllister (Director of the School of Music, University of Arizona), and moderated by me. The description in the conference program will read: "Panelists will discuss student assessment, current requirements of the National Association of Schools of Music and the National Council for Accreditation of Teacher Education, and the movement toward standards and accountability in higher education." This was the top-ranked issue by members who responded to my emails, and as the conference approaches, we will develop specific questions and points for discussion. I hope to see you all there!

3. Guest Commentary: Richard J. Colwell

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.

James Sherbon of the University of North Carolina at Greensboro and I started the measurement and evaluation newsletter for multiple reasons. We had some evidence that the music educators interested in assessment were randomly scattered across the country. There appeared to be a need for communication on issues critical to assessment in the field; this communication consisting of ideas to ponder, tentative solutions, news from other disciplines, and even an opportunity to review and critique assessments in and out of the discipline. We saw a relationship between policy and assessment and encouraged readers to submit material on policy issues. The newsletter was an extension of the sessions at the annual in-service conferences. We didn't want to limit the material to music educators writing for other music educators, the purpose of most MENC publications.

Music is related to culture and music is taught to improve the culture in the good society. There is too much fragmentation with our age of specialization. With the many reports on intellectual work, we wanted others to discuss the insights attained in their scholarly endeavors.

Research studies in music and in education have an assessment component and yet there appeared to be little critical thought involved with the selection and use of research-based assessment and the related research included was seldom adequately critiqued. (I'm not alone as Karen Harris, editor of the *Journal of Educational Psychology* indicates that the quality of research in education has deteriorated—see *Education Gadfly* 6 (6) Feb 9, 2006, available at <http://www.fordhaminstitute.org/institute/gadfly/>).

Louise Pascale (2002) states that "our historians lack any critical analysis of the path taken by music educators. The picture painted is always a rosy one, implying that all decisions made were correct ones and the outcomes always positive. In fact there is almost a heralding of the decisions made and the path chosen by past leaders in the field of music education in the US." (Unpublished PhD dissertation, Lesley University. *Dispelling the Myth of the Non-singer*, p. 62).

Optimistically we thought the SRIG might serve as a clearinghouse for researchers, evaluators, and that policy gurus in the profession would look to SRIG members whenever assessment matters surfaced. In many areas we were overly optimistic. We failed to consider the political dimensions of assessment ventures. Nevertheless the newsletters were valued and used in many test and measurement classrooms. One example of our over-optimism should suffice. Educational Testing Service is responsible for the GRE in music education, for NAEP, for teacher certification tests, for advanced placement measures, and more. We thought close cooperation between ETS and the SRIG would be to everyone's advantage and we scheduled a

session at a national meeting that featured ETS personnel. Princeton took our idea of cooperation seriously and sent their top evaluator for each of their tests but few music educators attended this session. It was a great session for those of us passionate about assessment but at that point Jim and I realized the importance of a newsletter to communicate with the few and to avoid future embarrassment.

We were also not ready to limit our thinking to convention topics with known audience appeal. In our view that was not the primary purpose of the SRIG. The newsletter subscription base grew rapidly. We did not judge this to indicate a growing interest in assessment but to a plethora of graduate students who could not easily find material on priority assessment issues on their own campuses.

This report to the membership is an electronic conversation on issues that might be discussed in depth. I'm including some references in the text as one purpose of the SRIG is to call attention to interesting reports.

If we only had a few things right, the need for a continuation of a formal means of communicating among assessment scholars in music education is greater today than ever before. Our society is seemingly accountability driven; the public wants efficiency data on our refrigerators and our schools and if this accountability data will improve the product (refrigerators and students) by reporting the data, so much the better. The April 9, 2007 issue of *The New Republic* (Dennis Ross in *Statecraft, Ways and Means*) stated: "assessments are crucial for effective statecraft. Administrations must be governed by reality-based assessments, not faith-based assessments. They must see the world as it is and shape objectives and means...we do not need to cast aside ambitious objectives; we just have to see things clearly to understand how to change them."

Lee Cronbach (1982) suggested, in relation to assessment, he was like Descartes in 1637: "I encountered nothing so dubious that I could not draw from it some conclusion that was tolerably secure, if this were no more than the inference that it contained in it nothing that was certain" ("In praise of uncertainty: Standards for evaluation practice," in Peter Rossi [Ed.], *New Directions for Program Evaluation*, 15, 49-58).

Just as a shoemaker thinks of life in terms of shoes, I see music education through the lens of assessment—and I ask continually how do we truly know? I compare music education publications with multiple disciplines, including assessments of the World Bank and programs by the state and defense departments. For example, visual artists and art educators deal primarily with that which happens before art works are made. What comes afterwards is the responsibility of the humanities. Music educators, however, seem to have little concern with the humanities.

There are any number of issues I would like to suggest that knowledgeable assessment gurus should write about in a SRIG newsletter. I shall mention a number of topics (or sentences) about which I'm presently puzzling and end these remarks by making a few substantive concerns I have in state and national testing—that is, if I'm asked to write again.

There are topics where music educators are following other disciplines and adapting their rationales and strategies when this may not be appropriate, and there are other problem areas where we should be forging on alone but seemingly are holding back. There are also common practices that simply need to be publicly threshed out. Bret Smith has not limited the length of this historical introduction but it is better to have reasonably short articles and a frequent publication. It's OK to not explain every detail when communicating with knowledgeable individuals. That's why we enjoy dialogue. Bret will have to learn about his subscriber base. I'm personally tired of reading education publications, obviously not written for an educated audience, where the practice is to cite four references for statements such as "George Washington was the first U.S. president." Recently Debbie Meier and Diane Ravitch have carried on a daily internet dialogue, "Bridging Differences," of about 500 words (available at <http://blogs.edweek.org/edweek/Bridging-Differences/>). Perhaps it is as Max Planck once

observed that new ideas take hold only when the old professors retire or die.

Whenever possible we need meta-analysis studies or data from contrasting reports in any discussion of importance. When the teacher educators within AERA sought research data to justify methods, coursework in the arts and sciences, field experiences, foundation courses, and more, they found that none of these experiences could be justified on the basis of research (*Studying Teacher Education*, Cochran-Smith and Zeichner, 2005). Ramirez, et al. in the Nov 2006 issue of the *American Journal of Education*, 113 (1), found a lack of evidence to support the connection between student achievement and economic growth. Stanford researchers recently reported that ethanol is not likely to solve any of our energy problems, (*Christian Science Monitor*, April 19, 2007, <http://www.csmonitor.com/2007/0419/p17s01-stss.html>). One can find opposing research findings on Reading First, NCLB, medical interventions and on whether teachers certified by NBPTS are better able to increase student achievement (Debra Viadero, "Studies mixed on national certification for teachers", *Education Week*, March 7, 2007, <http://www.edweek.org/ew/articles/2007/03/07/26teach.h26.html>).

The SRIG has an important role, as writers in education and music education have tended to list only those studies supportive of their own viewpoint and/or findings. In *Mentoring Teachers Toward Excellence*, (Judith Shulman and Mistilina Sato, [Eds.], 2007, Jossey Bass), Maryellen Weimer states in "Enhancing scholarly work on teaching and learning" that teachers seldom read the professional literature and don't look at research when they do read (p. 20).

I wonder why data-knowledgeable individuals don't speak up about teacher turnover. In the 1960s the Bureau of Labor Statistics reported that the typical college graduate could expect to hold five jobs in a lifetime; today that person will hold four jobs by the age of thirty. Are we to expect that teachers are not representative of the general population? (Data from page 7 of *Educational Entrepreneurship: Realities, Challenges, Possibilities*, Cambridge: Harvard Education Press, 2006).

Our enhanced capacity to use knowledge for policy purposes doesn't seem to have increased our inclination to question claims presented to us. It is apparently easier to listen and swallow than to stop and ask: How could that be true? We are all gullible, at least some of the time and, of course, there is simply too much data coming at us.

The knowledge base for us to understand what is being done and what should be done extends back at least to the 1970s when there were many reasons for discontent with public education and assessment. Minimum competency testing caught on in reading and math (sounds a bit like NCLB?) and any educational reforms were largely symbolic. To comply with minimum competencies, standards were lowered, only an acceptable number of students failed, and reform was put off. In teacher education, the list of competencies extended to a hundred or more. Someone must have read Ralph Tyler's writings in the 1930s as performance testing based on identifiable tasks was adopted. Performance testing is again favored by some, but it's a narrow sample and time consuming. The idea of testing was as step forward but performance testing was inappropriate for measuring individual achievement. It takes from 5 to 20 tasks to obtain a reliable measure. Defining performance is an inherently political process. It is also an inherently governmental process. There are no correct answers when it comes to selecting performance measures.

A few ideas for future newsletters:

- What will we know after the evaluation that we do not know now?
- Did outcomes based education mark the end of academic education in America?
- Outcome evaluations are defined as those examining questions about the impact of interventions, while process evaluations examine how interventions are delivered and received, and how and why an intervention might or might not be effective. Is this how these

terms are used in music education?

- Evaluations can describe the magnitude or severity of a problem, its incidence or prevalence, the results of past efforts to ameliorate it, gaps in knowledge of the problem, and the problem's novelty. They are also used to justify cutbacks in programs.
- What do we measure in order to decide if our expectations have been realized? What does it mean to know something and what does it take to be able to use what we know effectively?
- The design of good assessment depends on a clear vision of what it means to know something, to be familiar with facts and information. According to a 2006 study by Mike Schmoker, (*Results Now: How We Can Achieve Unprecedented Improvements in Teaching and Learning*, Alexandria: ASCD), in a study of 1500 classroom observations, evidence of a clear learning objective was present only four percent of the time. Schmoker was experienced but the data indicate that even with systematic procedures, brief classroom observations are notoriously unreliable. Schmoker believes that the single greatest determinant of learning is not socioeconomic factors, race, or funding levels—it is instruction (p. 7). Evaluation can be meaningless when the purpose of instruction is fuzzy.
- We really don't know much about how evaluation is practiced—how it is done, how much, how well, what influence it has or what influenced the evaluators.
- Assessment for learning requires pre-testing as it is a process of seeking and interpreting evidence for use by learners and their teachers to identify where the learners are in their learning, their needs, and how best to get there. Little pre-testing or needs assessment data can be found in music education. Pre-testing is necessary to understand the local condition, to determine what can be measured and whether the situation is such that a credible assessment can be conducted. If students have little or no competence, the situation of course does not lend itself to pre- or post-testing. This pre-knowledge is necessary, as many interventions show no effect because the within-class difference (individual student) in competence is greater than any between-class difference.
- Learning might be growth, change, development, knowledge, cognition and acquisition according to recent literature. How do we use these terms? Learning is not to be equated with thinking, cognition, or problem solving. Learning has more to do with the appropriation of culture.
- Assessment practices do far more than provide information, they also shape people's understanding about what is important to learn, what learning is, and who learners are. I wonder if music assessment practices are that clear?
- Agreeing on a score does not mean that scorers have the same reasons for inferring the score from their observations. If this is true, which I believe, it throws doubt on the "reliability" reported on performance ratings in contests, concerts, and the classroom.
- The word validity is used carelessly. Judgments of validity are never a property of measuring instruments. Validity can only be judged when we examine assessment results in the context of a particular argument or narrative. Validity rests on the person who interprets data. What kind of validity is this? Internal, external, construct, statistical conclusion, or....
- Rubrics raise serious validity issues. They standardize scoring which increased the reliability but they also standardize whatever is to be rated. Standardized performance or creations are not necessarily good. Compliance with a rubric is not necessarily learning, according to Sharon Nichols and David Berliner (2007, *Collateral Damage: How High-stakes Testing Corrupts America's Schools*. Harvard Education Press, p. 129).
- The standards for educational and psychological testing need to be critiqued. An excellent beginning can be found in Noel Wilson's (2007) "A little less than valid: An essay

review", in volume 10 Number 5 of *Education Review*, available at <http://edrev.asu.edu/essays/v10n5index.html>.

- Standard error matters more than reliability.
- If outcomes require evidence of students' abilities to perform for a range of purposes and audiences, then a portfolio may be an appropriate assessment method. If you desire evidence of students' abilities to collect, analyze, and report information, then a project is the more appropriate assessment tool.
- In designing college instruction in assessment, it is foolish to think that students should be taught how to emulate the construction of standardized achievement tests. What is important is how to select assessments and then how to use and interpret the data.
- I find it alarming that FairTest and others fault assessment whenever a commercial company errs in scoring a few papers. If the public believes this to be a serious issue, assessment cannot fulfill its potential.
- Paul LeMahieu and Elizabeth Reilly (2004), in "Systems of coherence and resonance: Assessment for education and assessment of education" (in Mark Wilson [Ed.], *103rd Yearbook of the National Society for the Study of Education*, Part II) suggest that the ability to frame thought-provoking questions is the key to developing critical thinking, but this is not assessment—it is the essence of education. Recently, as reported in the *American School Board Journal*, 194 (4) May 2007, the Gardner museum in Boston was awarded \$750,000 to take a few students from two schools to the museum to determine gains in academics and critical thinking. There is no report in the journal that there were no gains in academics, but a claim for critical thinking because the students compared art works, talked about them, and commented on controversial art. This is basic to what all art educators do when they teach—see Nancy Lampert (2007) *Art Education*, 59 (5) 46-50, p 48.
- Diane Ravitch suggests that she wished she had a dollar for every time she has heard or read a paean to the importance of critical thinking skills. She obtained over a million hits on Google. For advocates, if students learn how to think, then it matters not at all if they never read great literature or study history. One cannot think critically without some knowledge of the topics under consideration. Lois Weiner (2007) criticizes the World Bank and international comparisons sponsored by OECD as one can read for literary experience, to be informed, or to perform a task. Influenced by the bank, comparisons are often only on reading to perform a task which may indicate the interest of business. (Lois Weiner, "NCLB, U.S. Education, and the World Bank: Neoliberalism comes home." In Christine Sleeter [Ed.], *Facing Accountability in Education: Democracy and Equity at Risk*. New York: Teacher College Press, 159-171)
- Malcolm Gladwell tells us in *Blink* that intuition is far superior to the critical thinking skills that so many educators prize (2005, Little, Brown and Co.).
- Nel Noddings in her forward to *Collateral Damage* states that high stakes testing is wrong—intellectually, morally, and practically. Any indicator to which high stakes are attached will be subject to corruption (p xi).
- Formative assessment is a friendly type of assessment but many teachers do not realize that formative assessment must lead to further learning and that the method of instruction usually has to be changed based on the results from formative assessment. This is a large topic. Paul Black and Dylan William have provided an excellent discussion base in "Classroom assessment is not (necessarily) formative assessment (and vice-versa)" in Mark Wilson [Ed.] (2004) "Towards coherence between classroom assessment and accountability", *103rd Yearbook of the National Society for the Study of Education*, Part I, chapter 9. Publishers now provide teachers in some subjects with formative assessments at a cost of from \$5 to \$75 which must change the character of formative assessment. Without a common metric to track achievement across grades and within grades, the value of embedded instructional assessment will be limited (embedded often discussed along with results from any

formative assessment).

From the editor: Part 2 of Richard Colwell's guest commentary will appear in the next issue.

4. Feature Article: Kelly A. Parkes, *Assessment In The Applied Studio*

Kelly A. Parkes is Assistant Professor of Music Education at Virginia Polytechnic in Blacksburg, VA. Her email is kparkes@vt.edu.

The study of assessment, measurement, and evaluation has become more important to music education researchers over the last two decades, especially as more accountability is expected in education in general. The large number and wide variety of presentations at this year's Symposium on Assessment in Music Education (Gainesville, FL) illustrate how far assessment is impacting music education as a discipline. Many published and unpublished research studies contribute to the dialogue about assessment and enrich our literature. Some inquiry examines the applied studio, and to a lesser degree the elements of assessment, evaluation, and measurement in the applied studio are also examined. A wide view of the applied studio will be taken in this interest article, with a focus on how assessment in the applied studio has been investigated. A full literature review will not be given here. Rather, current, interesting, and important research will be examined for findings, and suggestions for the directions of future inquiry will be offered and discussed.

Historical Perspectives

The applied studio (individual performance instruction) has long been established in Western art music, and most musicians since the classical period have learned their instrument or voice through the applied studio method. Colwell (1971) suggested that "it is often thought that music teachers are against systematic evaluation because they fear the exposure of poor teaching" (p. 41) and he added that a more important reason may be the "conservatory atmosphere." He states that "conservatories are trade schools; their emphasis is on the development of specific skills...these skills are constantly evaluated...lessons, recitals...are constantly filled with evaluation" (p. 41). It is true that faculty are making these evaluations, particularly in the 'conservatory atmosphere' but has the process of these evaluations been investigated for reliability and validity? Do these truly exist in the subjective discipline of applied studio music instruction?

Madsen (1965, 1988) has been critical of the applied studio teaching methods and indirectly skeptical of the evaluation methods also used. In 1965 he described the applied studio as the following:

"this sad state of teaching seems to be based on three basic tenets: rejection of another teacher's worth, religious dedication to one's own abstract ideas and methods (which probably could not stand the test of experimental investigation), and a firm belief that any student who does not produce from this inspired teaching is obviously untalented" (p. 63).

His criticism is based in the lack of reliable and valid methods. He also suggested in his senior researcher award acceptance address (Madsen, 1988) that "it would be wiser to consider current practices of musicians...the way that musicians pass on their applied art has not substantially changed in hundreds of years" (p. 134). If the procedures for assessment over the last two centuries are examined, some change is visible. The early development of the conservatoire in 1795 with the French model of Conservatoire National de Musique (Paris Conservatory) gave almost free tuition to any talented pupil (Stolba, 1998 p. 399). Musician success meant "concertizing and gaining public acclaim" (Stolba, 1998 p. 456). Student success in performances, recitals, and competitions, along with job placement, were the primary vehicles by which a judgment was made. Students who graduated from the Paris Conservatoire became influential in Western art music—former students such as Ravel,

Milhaud, Messiaen, Ibert, Franck, Boulanger, Debussy, Bizet, Berlioz, and Maurice André. Competition works were often composed to “test” outstanding performers, and the end-of-semester performance usually took place as a recital of some kind. Composers and performers seemed to work together creating products that are now seen as benchmarks in the literature, and themselves reached outstanding success in the profession of music making. Current music schools, departments, or programs may not have the resources for these procedures and outcomes today, at least when compared to private conservatories. This is not the forum to discuss the socio-political reasons that current students of the performance of Western art music may be different than students of a century and a half ago, but it is worth noting the differences between the conservatory system and the music school, department, or program model that the US now holds. Throughout the Romantic period many more music schools and conservatories were built in both Europe and the USA (Stolba, 1998 p. 399) and as we move into the 21st century, and the current political climate affects us, there has been a noticeable reduction in the number of performing positions available in Western art music. Yet we are graduating more performers than can be hired. Often these performers have received highly distinctive grades, or A grades, for most of their applied lesson evaluations. Without the real-life test of winning a job in a professional ensemble, is this a reasonable evaluation of their achievement as musicians? Is giving a grade to a musical performance in a jury setting a real-life or “authentic” assessment?

Mills (1987) suggested that in the Western classical tradition of assessment, a vocabulary has developed for the discussion of performance for the purpose of evaluation; some applied faculty prefer the verbal openness of the comment sheet at recital or jury time to convey feedback about a music performance. There is a long held oral tradition in the applied studio and the vocabulary is often instrument-specific. Can the comments given to students after a jury examination or recital performance be considered assessment, evaluation, measurement or simply feedback? What is being measured? Skills, progress, or interpretation of repertoire? Is a set of criteria being used and if so, is it the same for all faculty, and more importantly, all students? Researchers in the last two decades have indicated many possible answers as to the methodologies used in the applied studio. They have been investigated through a variety of angles. Kennell (2002) overviews the research about the applied studio in general, examining models, theories of instruction, evaluation of instruction, methodological issues, descriptive research, experimental research, and to a small degree, evaluation in the applied studio. Applied studio teaching methods are clearer than they were two decades ago and Kennell suggests “our professional knowledge of the studio instructional system is complex” (p. 252) yet there is still a great deal we don’t know about the applied studio, and more importantly, the assessment procedures used.

The Research

Assessment in the applied studio is sometimes interesting to those of us involved with assessment, measurement, and evaluation but is often overlooked by our colleagues in music performance for a number of reasons, but particularly due to a paucity of research about assessment in their literature. Kennell (2002) also points out that “the researcher who studies studio instruction is faced with the immediate challenge of straddling two competing musical communities: the world of the performer and the world of the researcher” (p. 243). It is not surprising that Brand (1992) likened the mysterious methodologies of the applied studio to voodoo. It can be frustrating to many researchers who walk this bridge between two communities but there is a clear need for it to be done. Some researchers have done so successfully, in the area of assessment, and their findings have revealed promising answers to some of the questions posed.

The effectiveness of applied teachers has been well examined in the past, and while these findings are interesting, assessment is not usually identified as a characteristic of a good lesson. The work of Abeles (1975, 1992) and Livingston and Murray (1992) identified components of applied instruction and identified skills that novice applied studio teachers may require respectively, but assessment was not included. Bauer and Berg (2001) revealed that the top influences on high school instrumental music teachers when assessing students were their own experience, colleagues, and professional development activities. Schmidt (1998) also recognized the importance of personalized definitions of “good teaching” and using assessment

was not included. The work of Duke (1999) identified teacher feedback as a specifically useful tool used in lessons given in the Suzuki school method. The term *feedback* is usually used in the literature to refer to instant reinforcement that occurs within short teaching frames whereas the term *assessment* is generally used synonymously with grading or summative evaluation. Duke and Simmons (2006) reveal that goals and expectations are elements that are prominent in the lessons given by internationally renowned artist-teachers. The connection between these goals and expectations and the assessment points for measurement or evaluation is not made by Duke and Simmons; however, it is clear that the expectation of the artist-teacher is that the student play in a lesson as if they are performing on stage, that is, to achieve “a high standard.” This type of feedback is conveyed to the student consistently and it is reasonable to assume the student knows this expectation continues to prevail in the jury or recital setting. Duke (2005) has also acknowledged the importance of assessment in the teaching and learning paradigm in classroom settings suggesting that “assessment drives instruction” (p. 65) and that “tests teach” (p. 78).

Musical performance is always assessed and judged, yet much of the literature focuses on ensemble assessments. The reliability and validity of national and state festival ratings scales have been examined in recent years and the work of Bergee and Westfall (2005) has shown stability in explaining extramusical influences on festival ratings. The literature that represents assessment of solo performance was reviewed by Zdzinski (1991) and his report provides a detailed overview of the measurement instruments available for solo instrumental performance. The use of instruments has been examined more prolifically at the high school level and the work of Saunders and Holahan (1997) has pioneered the use of criteria specific rating scales in high school instrumental performance. Several researchers have focused on the use of assessment tools for the college level in applied performance studios.

Applied Studio Assessment Research

The work of Bergee (1993, 1997, 2003) tested the use of specific criteria rating scales, or rubrics, in the college applied studio setting. His work supports the concept that the criteria do help the faculty applied studio judges grade more consistently in the jury setting and shows that they grade with more reliability. The use of a specific tool in the applied studio measurement process is innovative. It is not apparent, in current research literature, the exact methods being used by most applied faculty but anecdotal evidence points to the predominant use of the jury comment sheet. Bergee’s findings illustrate his ability to walk in Kennell’s “two worlds” and test valid assessment tools for measurement of student achievement in the applied studio; for this, he must be commended. Parkes (2007) found great resistance from applied faculty to the use of a criteria-specific performance rubric. The reasons for resistance were due to many factors, but the underlying cause was largely an overt unwillingness from the applied faculty to be involved with research. They did not want, nor did they perceive a need, to use a measurement instrument in an otherwise unchanged protocol of jury comment-sheet grading. Oberlander (2000) didn’t examine the use of specific measurement tools but she did investigate the grading procedures in general. She showed that the overwhelming majority of clarinet instructors in the Northern USA and Canada give grades based on effort and improvement. Oberlander recommended that a fixed criterion be used in determining final grades to gain a higher level of objectivity. She also suggested determining in advance what level should be reached in order to pass particular criteria, keeping a written record of each lesson, the assignment of a grade each lesson, and having final grading techniques involve a screen to maintain anonymity of the student.

These three researchers have found some answers to the assessment procedures currently being used and Bergee, in particular, has uncovered a reliable and valid tool for use in the jury exam, the primary summative evaluation in many applied faculty studios. It remains unclear how many applied faculty, on a national level, use such a measurement tool, or how many continue to give summative evaluations based on extra-musical criteria. If some applied faculty are reluctant to engage in educational research and if there is, as Kennell suggests, a challenge in moving between researchers and performers, how will we, as researchers, build the bridge to connect the current disjuncture?

Direction Of Research In The Future

As investigators, we have the ability to impact teaching and learning. Colwell (1990, p. 30) has described the outcomes of research as “better teaching and more successful learning” and there can be no doubt that further investigation into assessment practices of the applied studio setting will indeed facilitate these outcomes. Price (1999) agrees and suggests that “with more systematic inquiry into little-investigated areas such as applied music instruction, we will have better-informed practice and beliefs” (p. 292). If music researchers undertake this task, what then might be the most salient research questions that should be answered about assessment in the applied studio?

Firstly, the grading system itself could be examined. Gordon (1971) suggested that “where grades are awarded in music, they are reported in the form of A, B, C, D, or F. Not only does this five-point system fail to provide a precise interpretation but, more importantly, grades ... do not describe the criteria by which a student was evaluated” (p. 132). Given Gordon’s theoretical stance, it is likely he is critical of the distinction between tonal and rhythmic achievement, however, he makes an important observation. There is little to be inferred from a letter grade in the applied studio as we cannot know what criteria it represents. We need to know more about how applied faculty give their grades; overall grades, jury grades, recital grade, and continuation exam grades. Some potentially effective investigative techniques are qualitative in nature, for example, sitting with applied colleagues informally, and simply asking “what factors are important to you in giving grades?” Clearly, a nationally representative sample would allow for the most inferential information but as a start, we could individually bridge the current gap that exists between researchers and performers by opening a personal dialogue with our performance colleagues. A more quantitative approach could entail the detailed observation of jury exam grading and recital grading by asking applied faculty to share their current tools, (rubrics, checklists etc.) with us for testing. We may find that instruments generated by applied faculty do have reliability and validity and that they are worth being made widely available.

Documenting applied lesson syllabi may also further illuminate how grades are given. Possible research questions could be: Which percentage of the total applied grade is given to the jury, to improvement, to attendance, or to attitude? Are applied faculty using extra-musical skills and/or non-music skills such as attendance and attitude to create a grade that is supposed to represent executive musical performance skills? Applied faculty may be more open to sharing their current methods regarding assessment if we, as researchers, take a genuine interest in understanding what it is that they do. If, or when, the time comes that they would like assistance in delivering more reliable, less subjective assessments, we will be able to do so with a sense of the best direction to take.

Perhaps we can look to the standards given by the National Association of Schools of Music (2007-2008). They currently state the performance standards for Baccalaureate Degrees in Music as:

- a. Technical skills requisite for artistic self-expression in at least one major performance area at a level appropriate for the particular music concentration.
- b. An overview understanding of the repertory in their major performance area and the ability to perform from a cross-section of that repertory.
- c. The ability to read at sight with fluency demonstrating both general musicianship and, in the major performance area, a level of skill relevant to professional standards appropriate for the particular music concentration.
- d. Knowledge and skills sufficient to work as a leader and in collaboration on matters of musical interpretation. Rehearsal and conducting skills are required as appropriate to the particular music concentration.

e. Keyboard competency. (*NASM Handbook, 2007-2008, VIII B.1, p. 84*)

We might write, collaboratively with our applied faculty, performance standards suitable for our individual departments, and then help develop tools by which to measure such standards. Music teacher educators must account for how they train, assess, and license pre-service music teachers, and show how they fulfill the standards prescribed by the National Council for Accreditation of Teacher Education (NCATE). Perhaps our research could focus on how we can assist our performance colleagues to better measure the NASM performance standards and how they can be more clearly reported. The technical skills could certainly be more objectively measured if schools and departments could collectively agree on suitable benchmarks for their students. Varied repertoire and sight-reading skills are relatively easy to measure and report from the applied studio jury exam, particularly when a cumulative approach from freshman to senior level is taken. Leadership, rehearsal, conducting, and keyboard skills might also be measured and along the same paths as pre-service music education students'. Simply having 'taken the course' is not a valid measurement of the skills acquired and more transparent evaluation and reporting may alleviate the perceived inequities that exist.

Investigative research with applied performance students might also yield answers as to how we might better prepare them to be performers, for example, inquiry to find out if our students hold high self-efficacy in regard to being a musician or the influences that influence success or failure as performance students. Much of this inquiry has occurred with our pre-service music education students but perhaps we should examine the performance majors with a sharper eye, with a focus on the summative evaluation they receive as part of applied lessons. Results from their perspectives may assist our understanding of the applied studio assessment strategies.

While it is clear that there is not one research question but many, it will be the task of interested researchers to take the first steps. There are many paths still to be explored that could illustrate, illuminate, and clarify what is happening to assessment in the applied studio. Every musician, and every music researcher, has participated in the applied studio method regardless of school or program size or location. It is the main vehicle by which music performance is taught and will remain so for quite some time. It is important that we, as researchers, deliver a commitment to establishing better evaluation procedures in this arena, which may ultimately lead to better accountability within our larger university settings.

References

- Abeles, H. (1975). Student perceptions of characteristics of effective applied music instructors. *Journal of Research in Music Education, 23*(2) 147-154.
- Abeles, H., Goffi, J., & Levasseur, S. (1992). The components of effective applied instruction. *The Quarterly Journal of Music Teaching and Learning, 3*(2), 17-23.
- Bauer, W. I. & Berg, M. H. (2001). Influences on instrumental music teacher. *Bulletin of the Council for Research in Music Education, 150*, 53-66.
- Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education, 41*, 19-27.
- Bergee, M. J. (1997). Relationships among faculty, peer, and self-evaluations of applied performances. *Journal of Research in Music Education, 45*, 601-612.
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education, 51*, 137-150.
- Bergee, M. J. & Westfall, C. R. (2005). Stability of a model explaining selected extramusical influences on solo and small ensemble festival ratings. *Journal of Research in Music*

Education, 53(2), 358-374.

- Brand, M. (1992). Voodoo and the applied music studio. *The Quarterly Journal of Music Teaching and Learning*, 3(2), 3-4.
- Colwell, R. J. (1971). Musical Achievement: difficulties and directions in evaluation. In Symposium on Research in Musical Behavior. *Music Educators Journal*, 57, (No. 8), Music in Special Education, 34-43 + 79-83.
- Colwell, R. J. (1990). Research findings: shake well before using. *Music Educators Journal*, 77 (3), 29-34.
- Duke, R. A. (1999). Teacher and student behavior in Suzuki string lessons: Results from the International Research Symposium on Talent Education, *Journal of Research in Music Education*, 47 (4), 293-307.
- Duke, R. A. (2005). *Intelligent music teaching: Essays on the core principles of effective instruction*. Austin, TX: Learning and Behavior Resources.
- Duke, R. A. & Simmons, A. L. (2006). The nature of expertise: narrative description of 19 common elements observed in the lessons of three renowned artist-teachers. *Bulletin of the Council for Research in Music Education*, 170, 7-19.
- Gordon, E. (1971). *The psychology of music teaching*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Kennell, R. (2002). Systematic research in studio instruction in music. In R. Colwell & C. Richardson (Eds.), *The New Handbook of Research on Music Teaching and Learning* (pp. 243-256). New York, NY: Oxford University Press.
- Livingston, C. & Murray, J. (1992). The preparatory program: an effective vehicle for preparing studio teachers, *The Quarterly Journal of Music Teaching and Learning*, 3(2), 51-55.
- Madsen, C. K. (1965). Experimental research in applied music. *Music Educators Journal*, 51 (6), 63-64.
- Madsen, C. K. (1988). Senior Researcher Award Acceptance Address. *Journal of Research in Music Education*, 36 (3), 133-139.
- Mills, J. (1987). Assessment of solo musical performance: A preliminary study. *Bulletin of the Council for Research in Music Education*, 91, 119-125.
- National Association of Schools of Music. (November, 2006). *NASM Handbook, 2007-2008*. Retrieved April 30, 2007 from http://nasm.artsaccredit.org/site/docs/Handbook/NASM_HANDBOOK_2007-2008.4.25.pdf
- Oberlander, L. M. (2000). *A survey of grading procedures in applied clarinet at American and Canadian Colleges and Universities*. Unpublished doctoral dissertation, Arizona State University.
- Parkes, K. A. (in press). Effect of performance rubrics on college level applied studio faculty and student attitudes. In T. Brophy (Ed.), *Proceedings of the Integrating Curriculum, Theory, and Practice: A Symposium on Assessment in Music Education*. Chicago: GIA Publications, Inc.
- Price, H. E. (1999). Forum. *Journal of Research in Music Education*. 47 (4), 292.

Saunders, T. C. & Holahan, J. M (1997). Criteria specific rating scales in the evaluation of high school instrumental music. *Journal of Research in Music Education*, 45 (2), 259-272.

Schmidt, M. (1998). Defining "good" music teaching: Four student teachers' beliefs and practices. *Bulletin of the Council for Research in Music Education*, 138, 19-46.

Stolba, K. M. (1998). *The development of western music*. New York, NY: McGraw-Hill.

Zdzinski, S. F. (1991). Measurement of solo instrumental music performance: A review of the literature. *Bulletin of the Council for Research in Music Education*, 109, 47-58.

5. MENC Draft Statement on Assessment

All SRIG members are urged to read and carefully consider MENC's draft statement on assessment available at <http://www.menc.org/connect/surveys/position/assessment.html>

Although the comment deadline was June 1, it may be possible to do so at this late date. For more information, e-mail [Sue Rarus](mailto:SueR@menc.org), (SueR@menc.org) MENC Information Resources Director. Responses and discussion sent to me can appear in future newsletters.

6. 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.