

2021

Planning Assessment in Music Education: It's Not Risk Free

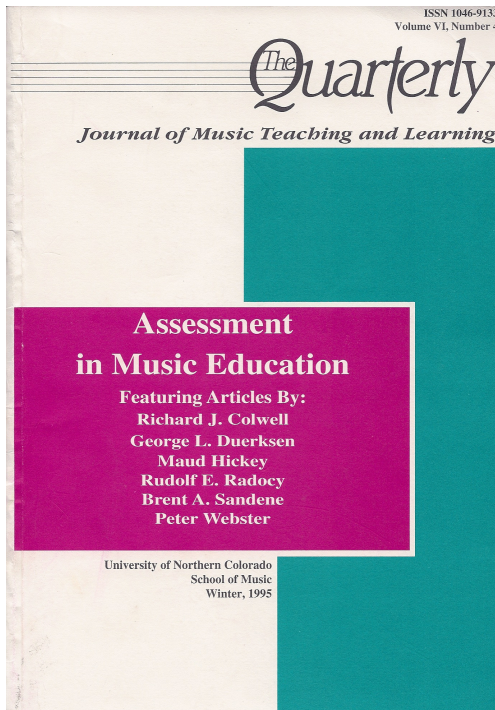
Rudolf E. Radocy
University of Kansas

Follow this and additional works at: <https://opencommons.uconn.edu/vrme>

Recommended Citation

Radocy, Rudolf E. (2021) "Planning Assessment in Music Education: It's Not Risk Free," *Visions of Research in Music Education*: Vol. 16 , Article 25.

Available at: <https://opencommons.uconn.edu/vrme/vol16/iss6/25>



Title: Planning Assessment in Music Education: It's Not Risk Free

Author(s): Rudolf E. Radocy

Source: Radocy, R. E. (1995, Winter). Planning assessment in music education: It's not risk free. *The Quarterly*, 6(4), pp. 19-25. (Reprinted with permission in *Visions of Research in Music Education*, 16(6), Autumn, 2010). Retrieved from <http://www-usr.rider.edu/~vrme/>

Visions of Research in Music Education is a fully refereed critical journal appearing exclusively on the Internet. Its publication is offered as a public service to the profession by the New Jersey Music Educators Association, the state affiliate of MENC: The National Association for Music Education. The publication of VRME is made possible through the facilities of Westminster Choir College of Rider University Princeton, New Jersey. Frank Abrahams is the senior editor. Jason D. Vodicka is editor of the *Quarterly* historical reprint series. Chad Keilman is the production coordinator. The *Quarterly Journal of Music Teaching and Learning* is reprinted with permission of Richard Colwell, who was senior consulting editor of the original series.

Planning Assessment In Music Education: It's Not Risk Free

By **Rudolf E. Radocy**
University of Kansas

Today's music educators must give considerable attention to planning assessment. Implementing national standards, local curricular standards and goals, and "outcomes" based elementary, secondary, and collegiate education requires evidence of effectiveness, preferably in the form of demonstrated student learning. Concerns for accountability, data-driven decision making, and management by objectives and the resulting demands for data are not new, of course. Many readers may recall the behavioral objectives movement of the 1960s and the accountability movement of the 1970s. In comparison with many other educators, music educators long have conducted assessment, especially performance assessment, that goes well beyond paper-and-pencil testing programs. In this article, I offer definitions, expound on assessment's inevitable subjectivity, describe authentic assessment, recount some psychometric traditions, and offer recommendations. Throughout, I maintain that one

must risk negative criticism and even embarrassment in the interest of conducting effective authentic assessment.

Definitions

The definitions which follow are my own, influenced by my experience and the experiences of uncited others. I recognize that others may define the terms differently, but I believe that these definitions are useful in the context of the discussions which follow.

Assessment is a process for rating or judging a person or event in accordance with specified or implied criteria or standards. It may be rather clear and reasonably objective, as in "assessing" whether or not a uniform or robe fits a student or assessing if a student has a sufficient number of credits to be eligible

for graduation. It may be somewhat subjective and open to alternative interpretations, as in assessing whether or not a high school student performs well enough to enter the band or choir, or whether a transferring college student has had courses or "real life" experiences that negate the need for a course normally required for graduation. It may be very subjective and largely without any "right" answer, as in determining which examples of all the world's music should be part of a music curriculum, or in assessing

Given the
inherent
subjectivity of
music and of
evaluation of
human actions, a
person should
accept that
subjectivity and
not hesitate to
make informed
judgments.

Rudolf E. Radocy is Professor of Music Education and Music Therapy at the University of Kansas. His scholarly interests include measurement and evaluation as well as the psychology and sociology of music.

I continue to believe that all measurement by humans, as well as of humans, is subjective.

who are the most significant composers who ever lived. The evidence from which one makes an assessment may be rather indirect, as it may be in the case of a paper-and-pencil test about musical instruments, or it may be performance-based, as in the case of having someone perform on the instruments or compose for them.

To *evaluate* is to place a value on some object or event, and the process for doing that is evaluation. One evaluates individuals regarding achievement, aptitude, attitude, and particular strengths and weaknesses. One evaluates groups in the interest of accountability, documenting instructional effectiveness, making policy, and evaluating particular projects (Boyle & Radocy, 1987, pp. 9-18). Presumably, evaluation is grounded in criteria, either explicit or implicit. While I will leave the task of distinguishing between evaluation and assessment to others, the realization that assessment and evaluation are something more than testing or measuring is important, as is the realization that standards considerably facilitate assessment and evaluation.

Standards are desired qualities, criteria, or characteristics. When we say that a musical performance must meet certain “standards,” we imply that it must have certain characteristics. For a performance by a sixth grade wind instrument soloist after one year of instruction, an appropriate standard might encompass playing correct pitches in their correct places at correct times, all at a musically appropriate tempo. With increasing proficiency, teachers might reasonably expect “correct” pitches to be within more rigorous intonation standards and rhythm figures to show some subtle variations in the interest of expression. Appropriate standards for a professional performance would go well beyond technical accuracy and would incorporate subtleties of interpretation and performance practice, as well as expectations for more sophisticated music literature. Standards for a student composition might specify a form, length, meter, style, and/or performance me-

dium. Standards provide a basis for assessment or evaluation. One takes risks in setting standards because differences regarding the propriety of particular standards for particular student populations are inevitable.

A *test* is a sample of behavior obtained under controlled conditions. Although the term often suggests a piece of paper containing questions, the test may be any procedure for sampling behavior. To *measure* is to assign numbers in accordance with rules; scoring a test by counting right answers and assigning a height by seeing where someone’s highest point reaches on an upright rod are obvious examples.¹ Sampling behavior and systematically assigning numbers are useful within assessment or evaluation, but assessment requires making judgments and interpretations — a process which inevitably is subjective, may be deemed “wrong,” and inevitably carries risks.

A *curriculum* is a set of experiences that promote education. This is rather broad; I intend the definition to recognize that much of what students learn is not an explicit part of any published curriculum guide or materials. Music curricula include the music, the assessments, teachers’ priorities, and incidental experiences that facilitate the teaching and learning of music. Teachers teach; students learn — the curriculum does neither. Understanding just what a particular music curriculum entails is essential in planning for assessment, especially with concern for “curricular alignment,” or matching assessments with what teachers allegedly teach and students allegedly learn.

Inherent Subjectivity

I have claimed that there is no such thing as objective measurement of human endeavors (Radocy, 1986). I continue to believe that all measurement by humans, as well as of humans, is subjective. “Subjective” implies individual feelings, an individual consciousness, an individual perception of how the world exists at the moment. “Objective” implies an external reality, an existence upon

Despite music educators' use of rather *authentic* tasks in evaluating performance and composition, most formal commentary in education seems to come from other fields.

which "rational" observers could agree. Indeed, judgment about whether the temperature is too hot, too cold, or just right involves highly individualistic reactions and preferences, while virtually anyone with normal vision and an appropriate observation point can agree regarding what an "objective" thermometer says. Surely it is easier for music educators to agree that a Beethoven symphony indeed is by Beethoven, or that an ensemble's instrumentation indeed is as a program says, than it is for them to agree that Beethoven's Ninth Symphony is his "greatest" or that the ensemble should have fewer of one instrument and more of another. Yet the decision to observe the temperature, look at the score's label, or count the instruments is itself subjective, albeit at a different level than evaluating the answer to the implied question. The multiple choice test, a paradigm of objectivity from the standpoint of observing a respondent's answer choices, is inherently subjective because the decisions to ask the particular questions of a particular group at a particular time are subjective. Furthermore, objective temperature as indicated by a thermometer and the "facts" of the ensemble's makeup or Beethoven's work are but part of what determines the inevitably subjective judgment of temperature comfort, the symphony's relative distinction, or the propriety of the ensemble's texture.

Subjectivity-objectivity is, I submit, a continuum rather than a dichotomy. The degree of subjectivity is greater in an essay examination than in a multiple choice examination; the evaluation of a university faculty member for tenure on the basis of what his or her publications say is more subjective than just counting the number. Some of the most important music learnings, including composition skills, performance interpretations beyond the printed score, and evaluating the aesthetic worth of a composition inevitably are highly subjective. *Music* is subjective, the result of the human brain's subjective recon-

struction of auditory signals. The signals, while "objective" and quite able to exist independently of any observer, are not music until the observer "subjectively" acts upon them.

Given the inherent subjectivity of music and of evaluation of human actions, a person should accept that subjectivity and not hesitate to make informed judgments. There is risk that one's criteria for evaluating a performance, the musicality of an arrangement, or the appropriateness of a harmonization may differ from the criteria that other knowledgeable persons would employ. Accuracy is a crucial part of performing and creating music; eliminating "wrong" notes is a major reason for rehearsal. An evaluator may risk overlooking some "wrong" notes in the process of evaluating a larger musical scope. An "informed" judgment may not be as informed as it could have been if the evaluator had spent more time researching relevant information. Yet, the person charged with assessment inevitably reaches a point where the risk entailed in making the assessment is justified. That risk may be especially justified in the context of an "authentic" assessment.

Authentic Assessment

"Authentic" assessment of formal educational outcomes essentially is an assessment procedure based on skills necessary for applying school-based learning in the world beyond the school. While many performances occur within the school, the public aspect of performance makes placing the results of teaching performance skills before an audience quite *authentic*. In principle, teachers may assess any part of music learning *authentically*. Of course, taking risks is necessary. Despite music educators' use of rather *authentic* tasks in evaluating performance and composition, most formal commentary in education seems to come from other fields.

Wiggins (1989) identifies common characteristics of authentic assessment. One, perhaps the most obvious, is that authentic as-

There are risks in authentic assessment, regardless of its thoroughness and complexity. What, after all, is *authentic*? What are the important *real world* behaviors for musically literate people?

assessments represent actual real-world performance. Students must engage in tasks that reflect ongoing practices; e. g., writers write intact essays and reports rather than completing grammatical exercises, and incipient scientists conduct experiments of their own design rather than filling in the blanks on a guide sheet. Musicians might arrange a song for a small ensemble rather than answer questions about playing ranges or completing notation exercises. The more mundane exercise aspects are incorporated within the holistic performance task.

Another characteristic of authentic assessment is that evaluation criteria are grounded in precise standards, standards of which students are well aware. Students not only know that an essay is due but also know stylistic, formal, and content criteria by which teachers will evaluate the essay. Music teachers would present a composition assignment with particular constraints regarding style, length, and performance medium. Teachers encourage students to engage in extensive self-assessment of their work.

Yet another characteristic is frequent public presentation of student work. Students share results of their learnings with peers, parents, teachers, and other interested parties. Public presentations are often oral, although public performance of music obviously fits this characteristic of authentic assessment.

Darling-Hammond, Ancess, and Falk (1995) extol the virtues of authentic assessment and present five case studies of its application in public schools. While the applications vary, one common component is that assessment procedures are an inextricable part of the curriculum, not only as it exists on paper but as it exists in daily pedagogical practice. Assessment is far more than gathering data a few times during the school year. The authors state:

This is one salient characteristic of an 'authentic assessment': it is designed to provide the student with a genuine rather than a contrived learning experience that provides both the teacher and the student with opportunities to learn what the student can do. The demonstration of learning occurs in a situation that requires the application and production of knowledge rather than the mere recognition or reproduction of correct answers. Authentic assessments are also contextualized: that is, rather than assembling disconnected pieces of information, the tests are set in a meaningful context that provides connections between real-world experiences and school-based ideas. (pp. 3-4)

Writing in the context of assessing young children, Lee (1992) notes that alternative assessments² may be either nonstructured or structured. Nonstructured assessments include those which are part of a regular classroom routine. The portfolio is a collection of nonstructured assessments, albeit a carefully selected collection. Structured assessments are predetermined tasks specifically for assessment. Lee suggests that teachers must become more skilled at the use of subjective evaluation procedures.

Examples of authentic assessment in action accumulate. Morrow (1992) believes that New York's use of a portfolio procedure to assess artistic production, perception, and reflection builds a strong case for arts education by documenting rigorous creative thinking. O'Neil (1992) describes Vermont teachers evaluating fourth and eighth grade students' portfolios, which are collections of students' "best work," of mathematics and writing examples during intensive half-day sessions. Kentucky implemented portfolios and performance assessments for fourth, eighth, and twelfth grade students. O'Neil notes that cost and time are possible negatives of the authentic assessment process, and that the "high stakes" nature of accountability for stu-

Tests... represent a long tradition of psychometric practice and custom. Breaking with tradition may carry some risk.

dent learning suggests a need to move slowly into performance-based assessment.

Schectman (1992) modified the assessment center method,³ used in the business world to evaluate management trainees and other personnel, to create a group assessment model for predicting the on-the-job success of a sample of Israeli teachers. Trained faculty observers watched newly admitted education majors engaging in intensive discussions and rated the students' oral communication, human relationships, leadership, and overall skill. Later, faculty evaluated students' teaching skills; principals eventually evaluated the 78 graduates who became teachers. In all cases, there was a significant ($p < .01$) correlation between scores obtained from the prediction procedure and the various criterion scores.

There are risks in authentic assessment, regardless of its thoroughness and complexity. What, after all, is *authentic*? What are the important *real world* behaviors for musically literate people? To what extent should we base assessment of music education on public performance when much of the public is involved passively with music? How should we make musical analysis and criticism and the results of creative thinking public? The much-heralded *National Standards for Arts Education* (Consortium of National Arts Education Associations, 1994) vary in their specificity; evaluative interpretations of the standards may vary considerably. For example, music content standard 3a for grades 5-8 (p. 43) says "Students...improvise simple harmonic accompaniments." Creating accompaniments is a useful musical skill, and the standard certainly is worthy. In assessing the created accompaniments, however, the teacher must interpret the standard in the context of a particular instructional situation. Is "simple" I-V-I? I-IV-V-I? Are block chords sufficient? Is an arpeggiated accompaniment desirable? Does the simplicity imply a slow tempo? What will be accompanied, and on

what medium? Diverse answers to such questions exist, and, despite any consensus among a group of music educators, other music educators may criticize the answers and the resulting assessment procedures negatively. In the interest of assessing the *authentic* musical skill of creating accompaniments, the music educator who incorporates accompaniment creation into the curriculum probably will withstand the criticism or modify assessment procedures.

A Long Tradition

Tests, particularly paper-and-pencil tests, represent a long tradition of psychometric practice and custom. Breaking with tradition may carry some risk. Considerable expertise exists regarding item construction, enhancing test reliability and validity, and standardizing tests to make the scores interpretable in relation to a designated population. Standardized measures of musical aptitude and achievement enable teachers who agree with the tests' underlying philosophies to obtain reliable data for assessing their students in accordance with those philosophies. For example, someone who believes in the importance of Gordon's concept of audiation has the Gordon family of carefully developed measures available. The Colwell battery is available for someone who interprets musical achievement as answering questions rooted in the consensus opinions of a panel of music educators that convened about a quarter of a century ago.

Locally developed tests often do not have the extensive refinement characterizing published standardized measures. Yet the curricular alignment of a locally developed test with a program's educational goals may justify the risk of a lack of psychometric elegance.

In particular, validity may be a problem for performance-based assessments and other *authentic* measures. Moss (1992) suggests that the traditional three-fold concept of validity (content, criterion-related, and construct validity) is evolving into a concept that

Given all of the risks, how do we create and implement authentic assessments? How can we avoid shoddy work in the name of risktaking and yet apply assessment procedures for the benefit of music education?

construct validity is of paramount importance. Construct validity essentially justifies the interpretation of the test⁴ because the test purportedly represents some underlying theory. Furthermore, technical excellence may not be sufficient. Moss insists that the social consequences of a test's use must be considered within the overall validity. A test with a disproportionate failure rate among members of a particular subpopulation may be invalid due to undesirable social consequences. This is in accordance with Messick's (1975) earlier argument that the decision to use a test requires evaluating the potential consequences of a test's use as well as evaluating its technical excellence in the proposed context.

Writing in the context of reforming teacher education, Katz and Rath (1992) describe a dilemma between using specific or global assessment criteria. Specific criteria communicate more clearly regarding just what teachers expect. Global criteria allow more flexible judgments, taking into account particular contexts. Global criteria invite more criticism as being "subjective" and "unfair." The authors suggest that teachers could employ both types of criteria, with special reviews to resolve evaluative disparities.

Recommendations

Given all of the risks, how do we create and implement authentic assessments? How can we avoid shoddy work in the name of risktaking and yet apply assessment procedures for the benefit of music education? As with much of life, balance is necessary. Some recommendations include the following:

Assessment procedures should exist for a definitive purpose. One should not listen to individual students play or sing their parts, administer a published standardized test, have a quiz in a music class, or require students to assemble a portfolio of compositions or recordings simply to kill time, scare the

students, have a record of evaluation, or look busy. One should assess to obtain information regarding what students are learning about music and what they can do musically. To decide to implement a portfolio-based assessment or any other assessment procedure prior to deciding what information is necessary and how it will be used is inappropriate.

Given the establishment of particular goals for a particular music education program, the teacher/evaluator/assessor should employ multiple forms of assessment. One should not limit assessment to standardized measures, classroom achievement paper-and-pencil tests, auditions, individual and collective public performances, creative activities, or any other means for assessment. A variety of information, collected for clear and appropriate reasons, will enable a more complete assessment and understanding of what is happening as a result of instructional activity. One may give attention to details as well as global aspects of learning.

Standard psychometric practices remain appropriate. The concepts of reliability and validity remain viable despite changes in thinking regarding means for estimating them. Relations among items of a test are vital to the test's overall quality: Item discrimination and difficulty indexes provide guidance for item selection and interpretation. Descriptive statistics for central tendency and variability provide information about a group's performance. While norm-referenced standardized tests may be inappropriate as a major means of assessment, they are quite useful for ascertaining where particular students stand regarding their knowledge of particular skills that a testmaker deemed important.

Music educators should seek consensus in setting standards, yet not be afraid of disagreement. Choice of literature, appropriate learning sequences, the degree to which a

Standards are not immutable.... music educators may change standards in recognition of changing instructional priorities, student characteristics, and practical concerns. Standards are more effective when they contain clear suggestions for assessment...

student must “excel at” or “know” something, and age appropriateness for particular tasks are all legitimate matters for debate. Eventually, people with responsibility for assessment must make decisions based on that debate and live with them.

Standards are not immutable. Regardless of particular standards’ elegance, propriety for a particular program, and the endorsement of authorities, music educators may change standards in recognition of changing instructional priorities, student characteristics, and practical concerns. Standards are more effective when they contain clear suggestions for assessment, especially when there are descriptions of just what is evidence of the achievement for which the standard calls. Opinions regarding appropriate evidence may change.

Music educators should recognize that music education is part of a broader educational enterprise, all of which necessarily is concerned with assessment. Even though we may be ahead of many academic areas in assessment, we should be aware of continuing developments.

Notes

1. My definitions of “test” and “measure” are close respectively to those of Cronbach (1970, p. 26) and Stevens (1946; 1951).

2. The “alternative” in alternative assessment is an “alternative” to standardized testing, which “authentic” assessments very well may be.

3. According to Zedeck (1986), the assessment center method requires personnel to engage in exercises that are representative of real-world tasks appropriate for their likely careers. During an intensive three- or four-day session, multiple observers employ varied subjective and “objective” measures of the persons’ skills and ability to work together.

4. “Test” is in a broad context of a sample of behavior; it is not limited to paper-and-pencil measures.

References

- Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. New York: Schirmer Books.
- Consortium of National Arts Education Associations. (1994). *National Standards for Arts Education*. Reston, VA: Music Educators National Conference.
- Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.
- Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action*. New York: Teachers College Press.
- Katz, L. S., & Rath, J. (1992). Six dilemmas in teacher education. *Journal of Teacher Education*, 43, 376-385.
- Lee, F. Y. (1992). Alternative assessments. *Childhood Education*, 69, 72-73.
- Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, 30, 955-966.
- Morrow, S. D. (1992). Assessment in the arts: Portfolio/performance. *Childhood Education*, 69, 96M.
- Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62, 229-258.
- O’Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, 49, (8), 14-19.
- Radocy, R. E. (1986). On quantifying the uncountable in musical behavior. *Council for Research in Music Education*, 88, 22-31.
- Schectman, Z. (1992). A group assessment procedure as a predictor of on-the-job performance of teachers. *Journal of Applied Psychology*, 77, 383-387.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677-680.
- Stevens, S. S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1-49).
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46 (7), 141-147.
- Zedeck, S. (1986). A process analysis of the assessment center method. *Research in Organizational Behavior*, 8, 259-296.

