

Measurement of Instrumental and Vocal Undergraduate Performance Juries Using a Multidimensional Assessment Rubric

Charles R. Ciorba

Neal Y. Smith

Millikin University, Decatur, Illinois

Recent policy initiatives instituted by major accrediting bodies require the implementation of specific assessment tools to provide evidence of student achievement in a number of areas, including applied music study. The purpose of this research was to investigate the effectiveness of a multidimensional assessment rubric, which was administered to all students performing instrumental and vocal juries at a private Midwestern university during one semester ($N = 359$). Interjudge reliability coefficients indicated a moderate to high level of agreement among judges. Results also revealed that performance achievement was positively related to participants' year in school (freshman, sophomore, junior, and senior), which indicates that a multidimensional assessment rubric can effectively measure students' achievement in the area of solo music performance.

Keywords: *assessment; rubric; performance*

From the earliest elementary grades to the highest levels of our nation's graduate schools, schools increasingly are being held accountable for documenting student learning. This phenomenon presents a unique challenge in the area of university-level jury performance. Because performance evaluation often is seen as a highly subjective endeavor, those responsible for judging solo music performances may have difficulty supplying data in a measure that provides evidence of student achievement. According to Fiske (1983), many evaluators are unaware of how they determined their performance ratings. To alleviate such difficulty, it is imperative for faculty to establish understandable criteria and objectives when creating assessments for university-level jury performances.

Likert-type scales have been used to assess solo music performance in a number of studies (Abeles, 1973; Bergee, 1988, 1989, 1993; Horowitz, 1994; Jones, 1986;

Address correspondence to Dr. Neal Y. Smith, Millikin University, Decatur, 1184 W. Main Street, Decatur, IL 62522; e-mail: nsmith@millikin.edu.

Winter, 1993; Zdzinski & Barnes, 2002). According to Saunders and Holahan (1997), Likert-type scales often can display a high degree of reliability, but their validity may remain uncertain. Specifically, decisions based on a Likert-type scale reflect an adjudicator's level of agreement with a general statement concerning a student's level of performance. Such decisions may not offer a consistent realization of performance achievement.

In contrast, criteria-specific rating scales use objective statements to provide written criteria explaining performance ability. They have been used by Azzara (1993), Levinowitz (1989), Rutkowski, (1990), Saunders (1990), and Smith (2006) to assess various levels of music achievement among elementary and middle school children. Saunders and Holahan (1997) created and administered a criteria-specific evaluation form to audition prospective students ($N = 926$) for the Connecticut High School All-State Band. Results indicated that the measure was able to assess accurately specific levels of performance ability with a high degree of internal reliability. Results of research by Abeles (1973) and Bergee (1987) suggested that valid solo performance rating scales could be developed using the facet-factorial approach, which provides a means of developing specific criteria for assessing music performance.

Whereas Likert-type and criteria-specific scales involve the quantitative assessment of several performance categories, holistic assessment awards a single score or grade to reflect an overall music performance. When examining judge-group differences in the assessment of solo trumpet performances, Fiske (1975) discovered significant correlations between the subscales and their composite scores. These results prompted Fiske to advocate the use of a holistic grading system when assessing music performances. Mills (1991) used multiple regression analysis to ascertain the degree to which holistic scoring can be explained by means of a segmented marking system. Mills used the composite scores from 12 bipolar constructs as the independent variable and a set of rank-ordered holistic scores as the dependent variable. The bipolar constructs combined to account for 71% of the variance in the holistic scores, which indicated that both approaches possess many commonalities.

Recently, rubrics have been used for the purpose of assessment in higher education to measure student progress in a uniform and empirical manner while providing clear learning expectations for the students. Kieke, Moro, and Gort (2007) described the process through which one university created a system of shared rubrics in the development of a university-wide assessment plan. By developing a set of general competencies, major programs were able to create a series of shared rubrics, each with a clearly defined set of outcomes to measure student learning. Bobbette (1999) outlined the development of a rubric designed to assess the portfolios of preservice teachers. Rubrics also have been used to assess learning outcomes in the areas of political science (Hutchins, 2003) and information literacy (Franks, 2003).

In the area of music assessment research, rubrics have been used to measure (a) the effect of keyboard ownership on music performance (Price, 2007), (b) the influence of certain variables on festival ratings (Bergee & Platt, 2003), and (c) the use of

imagery in the cognitive process to improve musical expression (Woody, 2006). Norris and Borst (2007) compared the reliabilities of a Likert-type rating form with a rubric when adjudicating choral festival performances. The authors reported that the rubric, with its clear performance descriptors, provided a more appropriate format for assessment. According to Asmus (1999), rubrics provide specific advantages when used to assess music performances. First, adjudicators are provided with clear descriptors outlining graduated levels of performance achievement. Second, performers are provided with (a) specific feedback concerning their performance and (b) useful information needed to improve future performances.

With the growing number of research studies using rubrics at various educational levels, a need exists to investigate the effectiveness of a multidimensional assessment rubric designed to measure student achievement in all areas of instrumental and vocal performance at the university level. By combining several performance dimensions into one standardized multidimensional assessment rubric, an assessment rubric can be used to assess multiple aspects of music performance, thus providing comprehensive evidence of student achievement. The purpose of this study was to create and test such a rubric. The following research questions guided this study:

Research Question 1: What is the interjudge reliability of a multidimensional assessment rubric when used for all the various performance areas (brass, woodwinds, strings, voice, piano, guitar, and percussion)?

Research Question 2: Is a multidimensional assessment rubric effective at describing differences in performance achievement among students at various stages of music study?

Research Question 3: What relationships exist between scores derived from a multidimensional assessment rubric and a holistic grading system?

Method

Instrument

Prior to the study, a faculty panel at a small Midwestern university met over a 6-month period to develop an assessment rubric that could be applied across all instrumental and vocal performance areas. Their intention was to create an assessment tool that was able to measure performance achievement in a consistent manner throughout the course of a student's undergraduate career. The panel comprised four experienced university faculty members who possessed performance expertise in the areas of brass, woodwinds, keyboard, and voice. All participants had earned their doctoral degree and had between 5 and 20 years of experience teaching music at the university level.

The panel first identified the following common dimensions of music performance, which were shared across all instrumental and vocal areas: (a) musical elements, (b) command of instrument, and (c) presentation. Panelists then created descriptors

outlining various levels of achievement for each dimension. Each descriptor encapsulated the characteristics of any performance medium. These descriptors were revised and edited to create five graduated criteria for each dimension on an equidistant continuum that ranged from a basic to an advanced level of expertise in each area.

The rubric was piloted over two semesters under performance jury conditions to determine the practicality of using the measure in relation to the time constraints of the jury setting. This gave faculty judges an opportunity to apply the rubric to a large range of student performers. Initial data were analyzed, and comments from the judges were solicited concerning the clarity and usability of the scale. As a result, minor changes in the presentation of the rubric were made in preparation for this study.

Participants and Judges

Music students were drawn from the university to participate in this study ($N = 359$). The university enrolls approximately 2,400 traditional undergraduate students and typically has between 300 and 350 music majors enrolled in eight different degree programs. Participating students, 153 male and 206 female, were enrolled in applied music study during the spring semester of 2006.

Thirty-seven judges participated in this study. All were current faculty members, 20 full-time and 17 part-time, who had extensive experience teaching applied music at the university level. Sixteen judges had an earned doctorate in music at the time of the study, and the remaining judges possessed a master's degree in music or its equivalent. Judges were asked to evaluate students within their particular area of performance expertise.

Procedures

The final version of the rubric (see Figure 1) was used during the spring semester's performance juries. Each participant played approximately 10 minutes for a panel of two, three, or four faculty judges. Participants were asked to play various études, solo repertoire, scales, and sight-reading material depending on the requirements of their performance area and their current level of performance expertise. Specific materials for participants were selected several weeks before the juries in consultation with their applied teachers. Because semester grades in applied music are determined in large part by jury performances, participants tended to place considerable emphasis on their preparation and presentation of material.

Judges were aware of each student's identity, and in most cases, one of the judges was the student's applied teacher. Although awareness of a student's identity may influence a judge's assessment of the student's performance because of prior experience he or she may have had with the student, the diagnostic and pedagogical value gained by the judges viewing the performances superseded concerns about biases that judges may have brought to the assessment. This also provided an authentic setting for the students and judges by reflecting typical jury practices.

Figure 1
Multidimensional Assessment Rubric

Total	1 point	2 points	3 points	4 points	5 points
	Musical Elements				
<i>Proficiency with and accuracy of musical elements, including pitch, rhythm, text articulation, and score accuracy</i>	Accuracy of musical elements does not meet minimal expectations (many noticeable mistakes)	Accuracy of musical elements meets minimal competency (a few noticeable mistakes)	Accuracy of elements is achieved most of the time	Accuracy of elements is proficient and well established	Precise demonstration of musical elements is demonstrated
	Command of Instrument				
<i>Represents the student's ability to control his or her instrument with musical intent</i>	Command of instrument is below minimum expectations (demonstrates little technical control of instrument)	Command of instrument demonstrates potential for musical growth	Command of instrument has achieved a point where musical maturity is possible	Command of instrument is proficient	Command of instrument demonstrates potential for professional success
	Presentation				
<i>Ability to be expressive with regard to phrasing, inflection, and style, as well as collaborate musically with others</i>	Presentation demonstrates a lack of sensitivity to musical expression	Presentation demonstrates a potential for musical growth	Ability to present a musical performance has achieved a point where musical maturity is possible	Presentation demonstrates that the ability to perform musically is proficient	Ability to perform musically demonstrates potential for professional success

Following each performance, judges assigned a score for each scale dimension of the rubric and provided written comments and a summative grade. Summative grades were based on a holistic impression of the students' performance compared with the faculty expectations. To arrive at their determinations independently, judges were asked not to consult with one another when determining scores and grades. Scale scores and grades awarded by each judge were averaged together to provide an overall score for each scale dimension, a composite score, and a grade for each student.

Results

Twenty-eight panels of judges assessed 359 undergraduate jury performances. Eight of the judges' panels assessed samples of less than 10 performers. These panels subsequently were removed from the analysis due to their small numbers.

Interjudge reliabilities were calculated for all scale dimensions, composites, and grades using Cronbach's alpha (see Table 1). Results indicated a significant level of

Table 1
Interjudge Reliability Coefficients for Brass,
Percussion, Woodwind, Voice, Piano, and String Faculty Groups

Panel	Judge n	<i>n</i>	Musical Elements	Command of Instrument	Presentation	Composite	Grade
Brass 1	3	12	0.95	0.87	0.92	0.95	0.97
Brass 2	3	14	0.94	0.88	0.91	0.93	0.97
Woodwind 1	3	11	0.82	0.80	0.76	0.88	0.73
Woodwind 2	4	14	0.59	0.59	0.63	0.61	0.91
Woodwind 3	4	12	0.95	0.91	0.91	0.95	1.00
Woodwind 4	3	28	0.90	0.94	0.94	0.96	0.98
Guitar	2	18	0.91	0.84	0.84	0.95	0.90
Percussion	2	10	1.00	0.97	0.94	0.99	0.89
Piano	4	24	0.92	0.93	0.91	0.95	0.92
Strings 1	4	11	0.92	0.93	0.89	0.95	0.88
Strings 2	3	10	0.84	0.89	0.98	0.93	0.98
Voice 1	3	43	0.77	0.80	0.80	0.66	0.87
Voice 2	3	22	0.74	0.81	0.81	0.87	0.89
Voice 3	4	21	0.85	0.86	0.84	0.91	0.94
Voice 4	4	11	0.86	0.71	0.80	0.85	0.93
Voice 5	3	12	0.91	0.87	0.84	0.94	0.83
Voice 6	3	13	0.76	0.82	0.75	0.90	0.80
Voice 7	3	13	0.92	0.82	0.83	0.87	0.96
Voice 8	3	12	0.70	0.31	0.70	0.83	0.56
Voice 9	3	14	0.76	0.74	0.73	0.88	0.90

agreement among members of 18 of the 20 remaining panels ($N = 325$). Only one panel, Woodwind Group 2, exhibited low reliability ratings for every scale dimension. An item analysis revealed that reliability ratings for this particular woodwind panel increased substantially when the scores assigned by two adjunct faculty members were removed. An additional low reliability rating (.31) was discovered in the command dimension of one vocal panel (Voice Group 8). Reliability also rose substantially when the scores assigned by one adjunct faculty member were removed. With the exception of these anomalies, reliability coefficients for each scale dimension were above .70 (elements .70–1.0, command .71–.97, presentation .70–.98). Reliability coefficients for the composite scores ranged from .66 to .99, whereas reliability coefficients for grades ranged from .56 to 1.0.

Means and standard deviations were calculated for all scale dimensions, composites, and grades (see Table 2). Based on a 4-point scale, the overall mean score for grades was relatively high (3.31), reflecting a negative skew. Scale dimensions, which were based on a 5-point scale, were distributed normally. In addition, standard deviations were narrower for grades than they were for scale dimensions in most groups. This result indicated a lack of variance in grades compared to scale dimensions.

Table 2
Descriptive Statistics for Brass, Percussion,
Woodwind, Voice, Piano, and String Faculty Groups

Panel		Musical Elements	Command of Instrument	Presentation	Composite	Grade
(Range)		(0-5)	(0-5)	(0-5)	(0-15)	(0-4)
Brass 1	<i>M</i>	2.75	2.67	2.72	8.14	3.20
	<i>SD</i>	1.06	1.01	1.01	2.39	1.99
Brass 2	<i>M</i>	2.76	2.69	2.69	8.14	3.34
	<i>SD</i>	0.90	0.78	0.88	2.54	0.40
Woodwind 1	<i>M</i>	3.40	3.24	3.18	9.82	3.91
	<i>SD</i>	0.51	0.58	0.52	1.53	0.14
Woodwind 2	<i>M</i>	2.80	3.24	2.54	7.75	3.66
	<i>SD</i>	0.40	0.52	0.40	1.07	0.34
Woodwind 3	<i>M</i>	3.58	3.50	3.75	10.83	3.99
	<i>SD</i>	0.87	0.83	0.78	2.47	0.03
Woodwind 4	<i>M</i>	3.36	1.12	3.18	9.61	3.70
	<i>SD</i>	0.95	0.32	1.01	2.67	0.60
Guitar	<i>M</i>	2.72	2.94	2.94	8.61	3.55
	<i>SD</i>	0.79	0.80	0.75	2.20	0.43
Percussion	<i>M</i>	3.18	2.59	2.73	8.50	3.65
	<i>SD</i>	0.60	0.92	0.82	2.00	0.43
Piano	<i>M</i>	2.94	2.83	2.66	8.43	3.38
	<i>SD</i>	0.70	0.65	0.64	1.89	0.37
Strings 1	<i>M</i>	3.08	3.05	3.09	9.22	3.76
	<i>SD</i>	0.69	0.80	0.76	2.20	0.23
Strings 2	<i>M</i>	2.40	2.10	2.07	6.57	3.41
	<i>SD</i>	0.72	0.78	0.68	2.14	0.78
Voice 1	<i>M</i>	2.54	2.48	2.49	8.05	3.04
	<i>SD</i>	0.76	0.75	0.84	4.26	0.58
Voice 2	<i>M</i>	2.54	2.39	2.52	7.36	3.23
	<i>SD</i>	0.69	0.66	0.83	2.04	0.44
Voice 3	<i>M</i>	2.55	2.36	2.27	7.18	3.17
	<i>SD</i>	0.62	0.65	0.71	1.79	0.56
Voice 4	<i>M</i>	2.34	2.55	2.52	7.41	2.65
	<i>SD</i>	0.80	0.53	0.66	1.74	0.70
Voice 5	<i>M</i>	2.58	2.22	2.50	7.31	3.26
	<i>SD</i>	0.74	0.73	0.69	2.09	0.29
Voice 6	<i>M</i>	2.56	2.41	2.28	7.26	3.14
	<i>SD</i>	0.69	0.56	0.57	1.64	0.48
Voice 7	<i>M</i>	2.95	2.92	2.87	8.85	3.10
	<i>SD</i>	1.02	0.75	0.81	2.29	0.79
Voice 8	<i>M</i>	2.42	2.00	2.19	6.61	3.08
	<i>SD</i>	0.41	0.35	0.36	0.93	0.30
Voice 9	<i>M</i>	3.17	2.80	2.90	8.88	3.27
	<i>SD</i>	0.76	0.66	0.70	1.89	0.52
Total	<i>M</i>	2.77	2.60	2.60	8.00	3.31
	<i>SD</i>	0.77	0.74	0.79	2.29	0.59

A one-way multivariate analysis of variance (MANOVA) was conducted to determine whether there were significant differences among scale scores and grades by participants' year in school. Results indicated a significant difference in scores by year, Wilks's $\Lambda = .75$, $F(6, 704) = 18.33$, $p < .01$. Table 3 contains the means and the standard deviations of scale scores and grades by year in school.

Analyses of variance (ANOVAs) on each dependent variable were conducted as follow-up tests to the MANOVA. Using the Bonferroni method, each ANOVA was tested at the .025 level. The ANOVA on the scale scores was significant, $F(3, 353) = 25.27$, $p < .01$, whereas the ANOVA on grades by year was nonsignificant, $F(3, 353) = .95$, $p = .42$.

Pearson correlations among scale dimensions, composites, and grades also were calculated (see Table 4). Correlations among scale dimensions and composites were found to be high (.81–.89), which indicates a strong relationships among all four areas. Correlations among scale dimensions and grades were moderate (.64–.72), which indicate some independence in grades compared to scale dimensions.

Table 3
Means and Standard Deviations of Scale Scores and Grades by Year

Year	Scale (0–15)		Grade (0–4)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	7.01	1.39	3.29	0.59
2	7.89	2.45	3.30	0.60
3	8.53	1.99	3.29	0.57
4	9.85	2.28	3.43	0.59

Table 4
Pearson Correlation Matrix for Scale Dimensions, Composites, and Grades

	Musical Elements	Command of Instrument	Presentation	Composite
Grade	.72*	.65*	.64*	.66*
Composite	.89*	.89*	.89*	—
Present	.81*	.88*	—	
Command	.82*	—		

* $p < .01$.

Discussion

The purpose of this study was to investigate the effectiveness of a multidimensional assessment rubric when administered across all university-level instrumental and

vocal performance juries. With the exception of two anomalies in the woodwind and voice panels, interjudge reliability was moderate to high across all dimensions. Furthermore, internal reliabilities were consistent within each performing area. These findings are similar to those determined by Bergee (2003), who found stable reliability when examining the assessment of performance juries across various instrumental groups. Reliabilities for scale dimensions and composites compared favorably with the reliability of the holistic letter grade.

Means for all scale dimensions and composites reflected higher levels of student achievement according to participants' year in school, whereas mean scores for grades remained consistently high, regardless of year. These findings are clarified by the results of the two ANOVAs, which indicated that scores derived from the rubric were related significantly and positively to students' year in school, whereas letter grades assigned by the jury panels were not. This supports the argument that a multidimensional rubric can prove useful when assessing student achievement from one year to the next.

Given the initiatives instituted by a number of accrediting bodies requiring evidence of student growth, the application of a multidimensional assessment rubric may provide an appropriate means to gauge performance achievement over time. This is consistent with previous research by Norris and Borst (2007), who reported positive results when using a rubric adjudication form to assess choral festival performances.

The relatively moderate degree of correlation between scale dimensions and grades suggests that there are different factors that come into play when determining grades as opposed to scale scores. These factors may include the degree of perceived effort exhibited by the student, the difficulty of repertoire attempted, the progress made over the course of the semester, or others. Because grades tend to carry a certain amount of personal importance to students, having two means of providing feedback seems appropriate. Students at all levels, freshman through senior, could certainly receive a high grade for their performance provided they performed at a level suitable to their status. Concurrently, the grades assigned may not capture the development of performance skills that students may have made over the course of their college experience.

The high correlations among scale dimensions are similar to those found by Fiske (1975). He argued that this strong relationship among dimensions supported the proposition that student performance could be captured effectively through a holistic measure, such as a letter grade. This argument fails to take into account the possibility that there could be diagnostic value in knowing a student's specific strengths or weaknesses in his or her performance skills. And although the unique contribution of each score to the composite may be limited, the comparison of scores in different dimensions presents a profile of student achievement that can lead to plans for future instruction to address areas of weakness.

We recommend that further research include a longitudinal study to determine whether this particular rubric accurately measures performance achievement for the same group of participants over a 4-year period. To establish content validity, we

further recommend including the administration of this rubric across various student populations from different universities.

Because judges were able to view performers and were aware of their identities, future researchers may want to involve an aspect of blind assessment to control for biases that judges may have when assessing students with whom they have had past experiences. It also may prove beneficial to examine the relationships between data derived from this rubric and other variables such as students' grade point averages, standardized test scores (ACT, SAT), years of study, self-efficacy, and self-perception. Such information could be used to develop a model designed to predict music performance achievement. Given the relative scarcity of research investigating the nature of multidimensional rubrics used to assess music performances, we hope that these findings will add to the potential value that this tool may offer in various educational settings.

References

- Abeles, H. F. (1973). Development and validation of clarinet performance adjudication scale. *Journal of Research in Music Education*, 126, 246–255.
- Asmus, E. (1999). *Rubrics: Definitions, benefits, history and types*. Retrieved March 17, 2008, from <http://www.music.miami.edu/assessment/rubricsdef.html>
- Azzara, C. D. (1993). Audiation-based improvisation techniques and elementary instrumental students' music achievement. *Journal of Research in Music Education*, 41, 328–342.
- Bergee, M. J. (1987). An application of the facet-factorial approach to scale construction in the development of a rating scale for euphonium and tuba music performance. (Doctoral dissertation, University of Kansas). *Dissertation Abstracts International*, 49 (05), 1086A. (UMI No. 8813388).
- Bergee, M. J. (1988). The use of an objectively constructed rating scale for the evaluation of brass juries: A criterion-related study. *Missouri Journal of Research in Music Education*, 5(5), 6–25.
- Bergee, M. J. (1989). An objectively constructed rating scale for euphonium and tuba performance. *Dialogue in Instrumental Music Education*, 13(2), 65–81.
- 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. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51, 137–150.
- Bergee, M. J., & Platt, M. C. (2003). Influence of selected variables on solo and small-ensemble festival ratings. *Journal of Research in Music Education*, 51, 342–353.
- Bobbette, M. M. (1999). Portfolios in a preservice teacher field-based program: Evaluation of a rubric for performance assessment. *Education*, 119, 416–426.
- Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education*, 23, 186–196.
- Fiske, H. E. (1983). Judging musical performances: Method or madness? *Update: The Applications of Research in Music Education*, 1(3), 7–10.
- Franks, D. (2003). Using rubrics to assess information literacy attainment in a community college education class. In E. F. Avery (Ed.), *Assessing student learning outcomes for information literacy instruction in academic institutions* (pp. 132–147). Chicago: American Library Association.
- Horowitz, H. (1994). *The development of a rating scale for jazz guitar improvisation performance* (Doctoral dissertation, Columbia University Teachers College). *Dissertation Abstracts International*, 55, 3443A.

- Hutchins, E. O. (2003). Assessing student learning outcomes in political science classes. In E. F. Avery (Ed.), *Assessing student learning outcomes for information literacy instruction in academic institutions* (pp. 172–184). Chicago: American Library Association.
- Jones, H. (1986). *An application of the facet-factorial approach to scale construction in the development of a rating scale for high school solo vocal performance* (Doctoral dissertation, University of Oklahoma). *Dissertation Abstracts International*, 47, 1230A.
- Kieke, M. C., Moro, K., & Gort, A. S. (2007). The transformation to a learner-centered community as a result of university-wide assessment. *On the Horizon*, 15(2), 107–117.
- Levinowitz, L. M. (1989). An investigation of preschool children's comparative capability to sing songs with and without words. *Bulletin of the Council for Research in Music Education*, 100, 14–19.
- Mills, J. (1991). Assessing musical performance musically. *Educational Studies (03055698)*, 17(2), 173.
- Norris, C. E., & Borst, J. E. (2007). An examination of the reliabilities of two choral festival adjudication forms. *Journal of Research in Music Education*, 55, 237–251.
- Price, H. E. (2007). Effect of keyboard ownership on keyboard performance in a music fundamental course. *International Journal of Music Education*, 25, 49–54.
- Rutkowski, J. (1990). The measurement and evaluation of children's singing voice development. *The Quarterly Journal of Music Teaching and Learning*, 1(1&2), 81–95.
- Saunders, T. C. (1990). A preliminary investigation of the suitability of selected rating scales used to measure student music performance skills. *Missouri Journal of Research in Music Education*, 27, 15–29.
- Saunders, T. C., & Holahan, J. M. (1997). Criteria-specific rating scales in the evaluation of high school instrumental performance. *Journal of Research in Music Education*, 45, 259–272.
- Smith, N. Y. (2006). The effect of learning and playing songs by ear on the performance of middle school band students (Doctoral dissertation, University of Hartford). *Dissertation Abstracts International*, 47, 1230A.
- Winter, N. (1993). Music performance assessment: A study of the effects of training and experience on the criteria used by music examiners. *International Journal of Music Education*, 22, 34–38.
- Woody, R. H. (2006). Musicians' cognitive processing of imagery-based instructions for expressive performance. *Journal of Research in Music Education*, 54, 125–137.
- Zdzinski, S. F., & Barnes, G. V. (2002). Development and validation of a string performance rating scale. *Journal of Research in Music Education*, 50, 245–255.

Charles R. Ciorba is assistant professor of music education at Millikin University. His research interests include performance assessment, self-perception, and jazz improvisation achievement.

Neal Y. Smith is associate professor of music education at Millikin University. His research interests include technology and creativity, performance assessment, and instrumental pedagogy.

Submitted April 7, 2008; accepted January 16, 2009.