

Illinois State University

ISU ReD: Research and eData

Faculty Publications - Music

Music

9-2013

An Analysis of Middle/High School Band and Orchestra Festival Ratings

Phillip Hash

Illinois State University, pmhash@ilstu.edu

Follow this and additional works at: <https://ir.library.illinoisstate.edu/fpm>



Part of the [Music Education Commons](#)

Recommended Citation

Hash, Phillip, "An Analysis of Middle/High School Band and Orchestra Festival Ratings" (2013). *Faculty Publications - Music*. 9.

<https://ir.library.illinoisstate.edu/fpm/9>

This Article is brought to you for free and open access by the Music at ISU ReD: Research and eData. It has been accepted for inclusion in Faculty Publications - Music by an authorized administrator of ISU ReD: Research and eData. For more information, please contact ISUReD@ilstu.edu.

AN ANALYSIS OF MIDDLE/HIGH SCHOOL BAND AND ORCHESTRA FESTIVAL RATINGS

Phillip M. Hash

Abstract

The purpose of this study was to compare festival ratings among multiple instrumental ensemble types (bands and orchestras), grade levels (middle school and high school), and classifications (1-6). Data included individual and final ratings from 144 judges (108 concert-performance, 36 sight-reading) at 36 contest sites sponsored by the Virginia Band and Orchestra Directors Association (VBODA) in 2010. Research questions examined the distribution, reliability, and group differences of ratings by ensemble type (band vs. orchestra), age level (middle school vs. high school), and classification (1-6). The average final rating was 1.58 ($SD = .66$) and 91.5% ($n = 901$) of ensembles ($N = 985$) earned either a I/Superior or II/Excellent out of five possible ratings. Data indicated a high level of interrater reliability regardless of contest site, ensemble type, age level, or classification. Although final ratings differed significantly by age (middle school bands vs. high school bands), ensemble type (middle school bands vs. middle school orchestras), and classification (lower vs. higher), these results were probably due to performance quality rather than adjudicator bias, since interrater reliability remained consistent regardless of these variables. Findings from this study suggested a number of opportunities for increasing participation and revising contest procedures in festivals sponsored by the VBODA and other organizations.

An Analysis of Middle/High School Band and Orchestra Festival Ratings

Many school instrumental and vocal ensembles participate in an annual adjudicated festival or contest. These events provide both students and directors with valuable feedback that can improve performance and raise the standard of musicianship (Rohrer, 2002). Nonetheless, music educators often are apprehensive about subjecting their ensembles to such intense evaluation (Barnes & McCashin, 2005; Batey, 2002), probably because many administrators, parents, students, and directors believe that ratings reflect teacher and program quality (Boyle, 1992; Burnsed, Hinkle, & King, 1985; Conrad, 2003). Furthermore, the numerous factors that can influence adjudication (e.g., McPherson & Thompson, 1998) and the subjective nature of performance evaluation (Garman, Boyle, & DeCarbo, 1991) might cause stakeholders to wonder if these events are fair and objective. Organizations, therefore, should work to establish and demonstrate interrater reliability for the events they sponsor.

Several researchers have analyzed the ratings and reliability of large-group festivals (Brakel, 2006; Burnsed, Hinkle, & King, 1985; Garman, Boyle, & DeCarbo, 1991; Hash, in press; King & Burnsed, 2009; Latimer, Bergee, & Cohen, 2010), solo and ensemble contests (e.g., Bergee, 2007; Bergee & McWhirter, 2005), and college performance juries (e.g., Bergee, 2003; Ciorba & Smith, 2009). Additional authors have examined performance assessment in laboratory settings using recorded examples with inservice (e.g., Norris & Borst, 2007) and

Hash

preservice teachers (e.g., Kinney, 2009; Price & Chang, 2005) as evaluators. Findings from these studies identified a number of variables other than performance quality that might affect final ratings in various performance situations.

Factors related to adjudicators that may affect assessment include judges' training and experience (Hunter & Russ, 1996; Winter, 1993), familiarity with repertoire and medium (Brakel, 2006; Garman, Boyle, & Decarbo, 1991; Kinney, 2009), and desire to encourage students and directors (Boeckman, 2002). Variables associated with contest and adjudication procedures can include the type of evaluation form used (Ciorba & Smith, 2009; Norris & Borst, 2007), length of the contest day (Barnes & McCashin, 2005), size of the judging panel (Bergee, 2003; 2007; Fiske, 1983), and performance order (Bergee & McWhirter, 2005; Bergee & Platt, 2003; Flores & Ginsburgh, 1996). Factors related to ensembles might include repertoire difficulty (Baker, 2004; Brakel, 2006), group size (Killian, 1998, 1999, 2000; King & Burnsed, 2009; Rickles, 2006), and labels such as "beginning" vs. "high school" (Cavitt, 1997) or "concert band" vs. "wind ensemble" (Silvey, 2009). Variables related specifically to students and teachers that may affect ratings include conductor expressivity (Morrison, Price, Geiger, & Cornacchio, 2009), race of performer (Elliot, 1995/1996) or conductor (VanWeelden & McGee, 2007), and special circumstances such as the participation of exceptional learners in the ensemble (Cassidy & Sims, 1991; Cavitt, 2002).

With all of the factors that might influence adjudication (e.g., McPherson & Thompson, 1998), it is not surprising that interrater reliability among judges working in actual large-group contests tends to be rather inconsistent. Brakel (2006), for example, analyzed the reliability of band and orchestra contests sponsored by the Indiana School Music Association in 2002 and 2003 and found that internal consistency for individual judging panels ($N = 43$) ranged from an alpha of .44 to .94 with a mean of .82 in 2002, and .76 to .94 with a mean of .87 in 2003. Interrater correlation was generally acceptable for ensemble classification in 2002, but ranged from $r = -.12$ to $r = 1.0$ in 2003. Average reliability by ensemble type also varied widely in 2002 but improved the following year after the organization offered training in orchestra evaluation for non-string specialists.

Hash (in press) utilized multiple measures to analyze the interrater reliability of high school concert band festivals sponsored by the South Carolina Band Directors Association (SCBDA) from 2008-2010. Reliability varied widely depending on contest site and type of event (concert vs. sight-reading). Although all reliability coefficients were higher for sight-reading versus concert-performance, each rose above .80 for all sites combined except interrater correlation ($r_s = .75$) and average pairwise interrater agreement (.70) in the concert portion of the contest. The author affirmed the reliability of SCBDA festivals and stated that lower readings for r_s and average interrater agreement were likely due to factors related to statistical procedures rather than the quality of adjudication.

A wide range of reliability coefficients for final ratings ($r = .54 - .89$) and individual captions ($r = .27 - .84$) were also attained in orchestra festivals in Dade County, Florida, in 1983, 1986, 1987, 1989, and 1990 (Garman, Boyle, & DeCarbo, 1991). Burnsed, Hinkle, and King (1985), in addition, found that that some individual captions (tone, intonation, balance, and musical effect) differed significantly in selected band contest sites in North Carolina ($n = 1$) and Virginia ($n = 3$), but that caption scores and final ratings were so highly correlated as to represent a single global performance rating. Similar results have been attained in other studies of solo (Fiske, 1975) and large-group (King & Burnsed, 2009; Garman, Boyle, & DeCarbo, 1991; Latimer, Bergee, & Cohen, 2010) assessment, as well. The fact that caption scores and final ratings were so highly correlated calls into question the validity of individual caption scores and suggests that judges may grade these in such a way as to arrive at a predetermined final rating.

Although reliability may vary widely depending on a number of factors, ratings often fall within a restricted range with most ensembles earning a Division I or II. This trend has been reported for concert bands (e.g., Boeckman, 2002; Hash, in press), marching bands (King & Burnsed, 2009), solo and ensemble contests (Bergee & McWhirter, 2005; Bergee & Platt, 2003; Bergee & Westfall, 2005), and large-group sight-reading assessment (Orman, Yarbrough, Neill, & Whitaker, 2007; Yarbrough, Orman, & Neill, 2007). For extensive reviews of the literature related to performance evaluation, see Conrad (2003), Forbes (1994), and McPherson and Thompson (1998).

The purpose of this study was to compare festival ratings and interrater reliability among multiple instrumental ensemble types (bands and orchestras), grade levels (middle school and high school), and classifications (1-6) using the same statistical procedures. Previous studies of large-group evaluation sometimes have examined only one type of ensemble (e.g., Garman, Boyle, & DeCarbo, 1991) and/or grade level (e.g., Hash, in press). Authors, furthermore, have used a variety of statistics to analyze ratings and reliability including Cronbach's alpha, Pearson's r , repeated measures ANOVAs (e.g., Burnsed, Hinkle, & King, 1985), Kendall's Coefficient of Concordance (Latimer, Bergee, & Cohen, 2010), and intraclass correlation (Norris & Borst, 2007). Although reliability can be expressed in a number of ways, studies that analyze ensembles of varying characteristics using the same procedures will facilitate comparisons of the data. This study will add to the literature on large-group performance evaluation, provide data with which to compare other research, and address some of the concerns regarding the fairness of contests often raised by stakeholders in music education (e.g., Conrad, 2003).

Data included individual and final ratings from 144 judges (108 concert-performance, 36 sight-reading) at 36 contest sites sponsored by the Virginia Band and Orchestra Directors Association (VBODA) in 2010. The following questions guided this research: (1) What was the distribution of ratings among ensemble types (bands vs. orchestras), age levels (middle school vs. high school), and classifications (1-6)? (2) What was the reliability of ratings for individual judging panels, ensemble types, age levels, and classifications? (3) Did average final ratings differ between ensemble types, age levels, or classifications?

Method

Participants

I analyzed ratings for 985 middle school ($n = 498$) and high school ($n = 487$) bands ($n = 596$) and orchestras ($n = 389$) participating in VBODA district concert festivals held in 2010. In these events, three adjudicators evaluated prepared selections and one judge assessed sight-reading. Bands and orchestras entered the festival under one of two program options.

Option one required ensembles to prepare three compositions from the VBODA Selective Music List that they had not entered in the previous three (middle school) or four (high school) years. Full orchestras could choose two full orchestra pieces and one string orchestra selection. In order to adjust to the acoustical properties of the auditorium, each group began their performance with an additional warm-up piece, which need not have appeared on the Selective Music List. Bands played a march and orchestras performed either a full or string orchestra selection. This piece was followed by two of the three prepared compositions from the Selective List. The adjudicators chose the first piece at the judges' meeting on the day of the event and the director selected the second. The director also decided the performance order of the two concert pieces. The warm-up selection and two concert works from the VBODA Selective List constituted the adjudicated program. The final rating was an average of the ratings given by each of the three judges.

The guidelines for music selection under option two were identical to those prescribed under option one except that directors selected and prepared only two compositions from the Selective Music List and played them both. Full orchestras were allowed to perform one piece for full orchestra and one for string orchestra. Following their concert performance, ensembles selecting option two proceeded to the sight-reading room where they had up to seven minutes to prepare and then perform one unfamiliar piece chosen by a committee appointed by the VBODA President. During the allotted preparation time, the director and students could engage in any type of instructional activity except performing on a musical instrument. The final rating under option two equaled the average of the three individual judges' ratings given on the warm-up selection and the two concert selections, and the single rating awarded by the sight-reading adjudicator.

The difficulty of repertoire determined the classification under which an ensemble entered the contest, with each performing group graded on the lowest level of any composition chosen from the VBODA Selective Music List. For example, a group that performed selections from grades II, III, and IV were classified at the grade II level. Bands and orchestras entering the festival in grade I were not required to sight-read and only had to prepare and perform a warm-up and two concert selections. Such groups had the option of sight-reading for comments only without their performance in this event affecting their final rating.

The concert-performance and sight-reading judges evaluated ensembles using forms developed by the National Music Adjudication Coalition (n.d.). Judges graded each category with a letter grade (A-F) and determined a numerical rating (I-V) based on the distribution of grades. Categories for the concert performance included tone, intonation, technique, rhythm, balance, musicianship, and general factors (concert-performance) or utilization of preparatory time (sight-reading). A final rating of I/Superior was equivalent to the letter grade "A" and represented a superior interpretation and performance, technically and musically, of all three selections. A rating of II/Excellent was equivalent to the letter grade "B" and demonstrated an excellent interpretation and performance of all three selections, or a superior performance of one selection and an excellent rendition of two. A rating of III/Good was equivalent to the letter grade "C" and denoted a good interpretation and performance of all three selections, technically and musically, or a combination of performances of the three selections which would justify an overall rating of "Good." A rating of IV/Fair was equivalent to the letter grade "D" and represented a performance that approximated the technical and musical requirements of the music but was seriously lacking in its rendition. The rating of V/Poor was equivalent to the letter Grade "F" and signified a performance which was unacceptable technically or musically. Contest rules stated that the concert adjudicators could confer after hearing approximately half of the performing groups in any one classification but should not release their grading sheets until the end of each of the classification performances (VBODA, 2010).

Data Analysis

Data for this study included individual and final ratings assigned by 108 concert-performance and 36 sight-reading adjudicators from 36 judging panels working in 13 districts of the VBODA in 2010. I downloaded ratings from the VBODA web site, transferred them to a Microsoft Office 2007 Excel database, transposed the classifications and ratings from Roman to Arabic numerals, and entered them into SPSS 17.0 for statistical analysis. Data was unavailable for all ensembles in District 7, middle school bands in District 9, and high school orchestras in District 12. This analysis did not involve grades for individual captions (e.g., tone, musicianship) because this data was unavailable and probably had little effect on results, since judges appear to mark individual categories based on the final rating they intend to issue (Burnsed, Hinkle, & King, 1985; Fiske, 1975; Garman, Boyle, & DeCarbo, 1991; King & Burnsed, 2009). The Institutional Review Board at [withheld] reviewed and approved these procedures.

Procedures for this analysis followed those of Hash (in press), who used multiple measures to examine the ratings and reliability of high school band contests in South Carolina from 2008-2010. Descriptive and frequency data examined final ratings awarded for each contest site, ensemble type, classification, and age group. I measured interrater reliability for the concert-performance judges using three statistical procedures, each of which described a different aspect of this construct. Internal consistency was determined using Cronbach's alpha (α) with the ratings of individual judges treated as "items." This calculation indicated the

degree to which judges' ratings corresponded with one another (Adler & Clark, 2008). I also calculated two forms of interrater agreement (IRA)—the average percent of pairwise agreement (IRA_{pw}) between individual judges (Freelon, 2009) and the percent of agreements for all ratings combined (IRA_{co}). Calculating IRA_{co} involved dividing the total number of agreements within each performance (0, 2, or 3) by the total number of ratings issued. Unlike IRA_{pw} , combined interrater agreement (IRA_{co}) considered ratings within each performance without comparing specific judges' decisions, thereby measuring the reliability of adjudication panels as single units as well as the combined assessments of judges from different contest sites. IRA_{pw} , however, was useful for identifying panels with frequent disagreement between two or more judges, and for affirming the system of checks-and-balances created by three-member adjudication panels.

Combined IRA_{pw} and alpha simply equaled the average of these statistics from each contest site. Total IRA_{co} was calculated using the entire data set, since this measure could be considered separately from specific adjudication panels. A benchmark of .80 (Carmines & Zeller, 1979; Krippendorff, 2004) represented the minimum level for good reliability in this study.

Additional analysis using Kruskal-Wallis ANOVAs and post hoc Mann-Whitney *U* tests examined differences in mean ratings among classifications, ensemble types, and age levels to determine the level of consistency between these variables. Wilcoxon signed-rank tests also compared concert-performance and sight-reading scores by age-level and ensemble type among the groups that participated in both events ($N = 730$) for a rating. I also used Spearman correlations to analyze the relationships between classifications and final ratings. These comparisons involved non-parametric statistics because contest ratings represent ordinal data with no absolute value or distance between ranks (Bergee & Westfall, 2005; Phillips, 2008).

Results

Distribution of Ratings

Final ratings varied by site from 1.23 ($SD = .44$) to 2.27 ($SD = .83$) for a mean of 1.58 ($SD = .66$) for all ensembles ($N = 985$, see Table 1). Most groups (91.5%, $n = 901$) earned a final rating of I/Superior (50.6%, $n = 498$) or II/Excellent (40.9%, $n = 403$). Only 8.5% ($n = 84$) of ensembles earned a III/Good (8.0%, $n = 79$) or IV/Fair (0.5%, $n = 5$) and no bands or orchestras earned a V/Poor (see Table 2). Average final ratings and the percentage of groups earning a I/ Superior increased with each advancing classification among bands with the exception of bands in class 3 that earned 27.2% Superior ratings compared to ensembles in class 2 that earned 29.2% (see Table 3). This trend did not hold with orchestras, where class 3 groups earned a higher percentage of Superior ratings than ensembles in all other classifications except class 6 (see Table 4). Further analysis indicated moderately weak or weak correlations between grade level and final ratings for both bands ($r_s = -.32$, $p < .001$) and orchestras ($r_s = -.16$, $p = .001$).

Table 1. Ratings and Interrater Reliability by Site

| District | Site | N | Level (MS/HS) | Ens. Type (Band/Orch.) | M | SD | IRA _{PW} | IRA _{CO} | α |
|----------|------|-----|---------------|------------------------|------|------|-------------------|-------------------|----------|
| 1 | 1 | 23 | HS | Band | 1.96 | .88 | .87 | .88 | .95 |
| | 2 | 34 | MS | Band | 1.56 | .61 | .82 | .85 | .91 |
| | 3 | 43 | MS/HS | Orchestra | 1.77 | .72 | .64 | .80 | .87 |
| 2 | 1 | 21 | HS | Band | 1.57 | .75 | .71 | .87 | .85 |
| | 2 | 24 | HS | Orchestra | 1.58 | .50 | .64 | .82 | .80 |
| | 3 | 16 | MS | Band | 1.31 | .48 | .71 | .85 | .87 |
| | 4 | 24 | MS | Orchestra | 1.38 | .50 | .81 | .90 | .87 |
| 3 | 1 | 24 | MS/HS | Band | 1.61 | .58 | .78 | .89 | .87 |
| | 2 | 34 | MS/HS | Band | 1.41 | .61 | .78 | .89 | .93 |
| | 3 | 29 | MS/HS | Orchestra | 1.34 | .48 | .71 | .84 | .62 |
| 4 | 1 | 34 | MS/HS | Band | 1.85 | .71 | .64 | .81 | .88 |
| | 2 | 29 | MS/HS | Orchestra | 1.34 | .55 | .76 | .86 | .86 |
| 5 | 1 | 26 | MS/HS | Band/Orch.* | 2.27 | .83 | .69 | .85 | .93 |
| 6 | 1 | 58 | MS/HS | Band/Orch. | 1.74 | .66 | .67 | .83 | .90 |
| 8 | 1 | 55 | MS/HS | Band/Orch. | 1.47 | .63 | .75 | .87 | .93 |
| 9 | 1 | 23 | MS/HS | Orchestra | 1.43 | .66 | .77 | .88 | .86 |
| | 2 | 27 | MS/HS | Orchestra | 1.67 | .56 | .73 | .86 | .81 |
| | 3 | 28 | HS | Band | 1.68 | .61 | .83 | .92 | .95 |
| 10 | 1 | 20 | HS | Band | 1.45 | .61 | .80 | .90 | .91 |
| | 2 | 24 | MS | Band | 1.79 | .65 | .72 | .86 | .89 |
| | 3 | 30 | MS/HS | Orchestra | 1.27 | .45 | .63 | .80 | .44 |
| 11 | 1 | 21 | HS | Band | 1.52 | .60 | .81 | .90 | .90 |
| | 2 | 21 | MS | Band | 1.62 | .67 | .84 | .92 | .92 |
| | 3 | 40 | MS/HS | Orchestra | 1.25 | .44 | .78 | .89 | .75 |
| 12 | 1 | 26 | HS | Band | 1.42 | .64 | .82 | .91 | .93 |
| | 2 | 26 | MS | Band | 1.65 | .69 | .79 | .90 | .92 |
| | 3 | 30 | MS | Orchestra | 1.37 | .556 | .76 | .88 | .82 |
| 13 | 1 | 26 | MS/HS | Band | 1.62 | .70 | .69 | .83 | .92 |
| | 2 | 14 | MS/HS | Orchestra | 1.36 | .50 | .76 | .88 | .82 |
| 14 | 1 | 22 | MS/HS | Band/Orch.* | 1.96 | .79 | .76 | .88 | .92 |
| 15 | 1 | 23 | HS | Band | 1.52 | .59 | .86 | .93 | .94 |
| | 2 | 29 | MS | Band | 1.72 | .59 | .75 | .87 | .85 |
| | 3 | 37 | MS/HS | Orchestra | 1.65 | .72 | .71 | .86 | .90 |
| 16 | 1 | 19 | HS | Band | 1.90 | .66 | .82 | .91 | .95 |
| | 2 | 12 | MS | Band | 1.58 | .79 | .78 | .89 | .93 |
| | 3 | 13 | MS/HS | Orchestra | 1.23 | .44 | .79 | .90 | .81 |
| Comb. | | 985 | | | 1.58 | .66 | .76 | .87 | .87 |

*This site only included one orchestra.

Table 2. Final Ratings and Interrater Reliability by Age Level and Ensemble Type

| Level/ Ensemble | N | M | SD | Frequency | | | | | Interrater Agreement ^a |
|--------------------|-----|------|-----|-------------|--------------|---------------|--------------|-------------|--------------------------------------|
| | | | | I (n, %) | II (n, %) | III (n, %) | IV (n, %) | V (n, %) | |
| MS Band | 277 | 1.74 | .68 | 107 38.6 | 135 48.7 | 34 12.3 | 1 0.4 | 0 0.0 | .87 |
| MS Orch. | 221 | 1.46 | .62 | 134 60.6 | 74 33.5 | 12 5.4 | 1 0.5 | 0 0.0 | .86 |
| Combined | 498 | 1.62 | .67 | 241 48.4 | 209 42.0 | 46 9.2 | 2 0.4 | 0 0.0 | .86 |
| HS Band | 319 | 1.60 | .70 | 164 51.4 | 121 37.9 | 31 9.7 | 3 0.9 | 0 0.0 | .89 |
| HS Orch. | 168 | 1.46 | .52 | 93 55.4 | 73 43.5 | 2 1.2 | 0 0.0 | 0 0.0 | .85 |
| Combined | 487 | 1.55 | .65 | 261 53.6 | 194 39.8 | 33 6.8 | 3 0.6 | 0 0.0 | .88 |
| Total | 985 | 1.58 | .66 | 498 50.6 | 403 40.9 | 79 8.0 | 5 0.5 | 0 0.0 | .87 |

^a Represents combined interrater agreement (IRA_{co})

Of the ensembles participating in the sight-reading portion of the contest for a rating ($N = 730$), 77.8% ($n = 568$) earned a I/Superior, 19.7% ($n = 144$) received a II/Excellent, and 2.5% ($n = 18$), earned a III/Good. No bands or orchestras received a IV/Fair or a V/Poor rating. Mean ratings for sight-reading by age and ensemble type ranged from 1.06 ($SD = .24$) to 1.30 ($SD = .59$) and averaged 1.25 ($SD = .49$) (see Table 5).

Table 3. Band Final Ratings and Interrater Reliability by Classification

| Grade | N | M | SD | Frequency | | | | | Interrater Agreement ^a |
|-------|-----|------|-----|-------------|--------------|---------------|--------------|-------------|-----------------------------------|
| | | | | I (n, %) | II (n, %) | III (n, %) | IV (n, %) | V (n, %) | |
| 1 | 24 | 2.00 | .83 | 7 29.2 | 11 45.8 | 5 20.8 | 1 4.2 | 0 0.0 | .81 |
| 2 | 125 | 1.90 | .66 | 34 27.2 | 70 56.0 | 21 16.8 | 0 0.0 | 0 0.0 | .86 |
| 3 | 163 | 1.79 | .72 | 62 38.0 | 75 46.0 | 25 15.3 | 1 0.6 | 0 0.0 | .87 |
| 4 | 145 | 1.59 | .67 | 72 49.7 | 62 42.8 | 9 6.2 | 2 1.4 | 0 0.0 | .88 |
| 5 | 96 | 1.44 | .70 | 59 61.5 | 32 33.3 | 5 5.2 | 0 0.0 | 0 0.0 | .89 |
| 6 | 43 | 1.14 | .35 | 37 86.0 | 6 14.0 | 0 0.0 | 0 0.0 | 0 0.0 | .98 |
| Comb. | 596 | 1.67 | .69 | 271 45.5 | 256 43.0 | 65 10.9 | 4 0.7 | 0 0.0 | .87 |

^a Represents combined interrater agreement (IRA_{co})

Table 4. Orchestra Final Ratings and Interrater Reliability by Classification

| Grade | N | M | SD | Frequency | | | | | Interrater Agreement ^a |
|-------|-----|------|-----|-------------|--------------|---------------|--------------|-------------|-----------------------------------|
| | | | | I (n, %) | II (n, %) | III (n, %) | IV (n, %) | V (n, %) | |
| 1 | 24 | 1.54 | .56 | 12 50.0 | 11 45.8 | 1 4.2 | 0 0.0 | 0 0.0 | .81 |
| 2 | 84 | 1.65 | .70 | 39 46.4 | 36 42.9 | 8 9.5 | 1 1.2 | 0 0.0 | .84 |
| 3 | 104 | 1.39 | .55 | 66 63.5 | 35 33.7 | 3 2.9 | 0 0.0 | 0 0.0 | .86 |
| 4 | 83 | 1.46 | .53 | 46 55.4 | 36 43.4 | 1 1.2 | 0 0.0 | 0 0.0 | .86 |
| 5 | 58 | 1.47 | .54 | 32 55.2 | 25 43.1 | 1 1.7 | 0 0.0 | 0 0.0 | .84 |
| 6 | 36 | 1.11 | .33 | 32 88.9 | 4 11.1 | 0 0.0 | 0 0.0 | 0 0.0 | .91 |
| Comb. | 389 | 1.46 | .58 | 227 58.4 | 148 38.0 | 14 3.6 | 1 0.3 | 0 0.0 | .87 |

^a Represents combined interrater agreement (IRA_{co})

Table 5. Sight-Reading Ratings by Age Level and Ensemble Type

| Age Level | Ensemble | <i>N</i> | <i>M</i> | <i>SD</i> |
|---------------|------------|----------|----------|-----------|
| Middle School | Band | 201 | 1.44 | .57 |
| | Orch. | 105 | 1.30 | .59 |
| | Combined | 306 | 1.39 | .58 |
| High School | HS Band | 288 | 1.18 | .41 |
| | HS Orch. | 136 | 1.06 | .24 |
| | Combined | 424 | 1.14 | .37 |
| Combined | Bands | 489 | 1.29 | .50 |
| | Orchestras | 241 | 1.17 | .45 |
| Total | | 730 | 1.25 | .49 |

Interrater reliability

Average pairwise interrater agreement (IRA_{pw}) among the concert-performance judges varied by festival site from .54 to .87, for an average of .76. Combined percent of interrater agreement (IRA_{co}) considered ratings within each performance without comparing specific judges' decisions. This coefficient ranged from .80 to .93 between sites (see Table 1), .81 to .98 among band classifications (see Table 3), .81 to .91 between orchestra classifications (see Table 4), .86 to .88 by age level, and equaled .87 for all ensembles ($N = 985$) combined (see Table 2). Internal consistency as measured by Cronbach's alpha (\square) varied by site from .44 to .95, for an average of .87 (see Table 1).

None of these procedures considered the magnitude of the difference between individual judges' ratings. However, only five ensembles—all orchestras—received scores more than one rating apart from the concert-performance adjudicators. Furthermore, 63.9% ($n = 381$) of bands and 58.6% ($n = 228$) of orchestras earned identical ratings from the three judges for this event. Additional analysis indicated a high correlation between measures of IRA (IRA_{pw} vs. IRA_{co} , $r = .90$) but a moderately low relationship between these readings and Cronbach's alpha (IRA_{pw} vs. \square , $r = .47$; IRA_{co} vs. \square , $r = .46$).

Ratings Comparisons

A Kruskal-Wallis ANOVA found significant differences in the final ratings of middle school and high school bands and orchestras ($N = 985$, $df = 3$, $\square^2 = 30.5$, $p < .001$). A post hoc

series of Mann-Whitney *U* tests determined the significant pairwise differences among the various groups. I applied the conservative Bonferroni correction to control for the increased chance of Type I error that results from multiple comparisons. Accordingly, an alpha level of .008 was considered the threshold for statistical significance (i.e., alpha of .05/6 comparisons). Data indicated a significant difference in the final ratings by ensemble type at the middle school level, where orchestras outscored bands, and among age groups, where high school bands earned significantly higher ratings than middle school bands. There was no significant difference between the final ratings of middle school and high school orchestras or high school orchestras and bands (see table 6).

Table 6. Ratings Comparisons by Ensemble Type and Age Level

| Age/Ensemble | <i>N</i> | <i>U</i> | <i>p</i> ^a |
|-----------------------|----------|----------|-----------------------|
| MS Band vs. MS Orch. | 498 | 23466.5 | <.001 |
| HS Band vs. HS Orch. | 487 | 24615.5 | .096 |
| MS Band vs. HS Band | 596 | 38747.0 | .004 |
| MS Orch. vs. HS Orch. | 389 | 17986.0 | .543 |
| MS Band vs. HS Orch. | 445 | 18232.0 | <.001 |
| MS Orch. vs. HS Band | 540 | 31524.5 | .018 |

^a Bonferroni correction for multiple comparisons ($\alpha = .008$)

An analysis of sight-reading ratings indicated that high school groups significantly outscored their middle school counterparts ($U = 51063.0, p < .001$) and that orchestras scored significantly higher than bands ($U = 51358.5, p < .001$). A Wilcoxon signed ranks test, furthermore, indicated that sight-reading ratings ($N = 730, M = 1.25$) significantly exceeded concert-performance ratings ($N = 730, M = 1.58$) ($Z = -12.54, p < .001$) among ensembles assessed in both events.

Kruskal-Wallis ANOVAs also found significant differences in the mean final ratings of different classifications for both bands ($N = 596, df = 5, \eta^2 = 65.49, p < .001$) and orchestras ($N = 389, df = 5, \eta^2 = 22.76, p < .001$). I, therefore, conducted two post hoc analyses using a series of Mann-Whitney *U* tests to find the significant differences among these groups. Using the Bonferroni correction, an alpha level of .003 was considered the threshold for statistical significance (i.e., alpha of .05/15 comparisons). The first post hoc analysis found that bands in class 6 received ratings significantly higher than bands in all other classifications. Furthermore, class 5 bands scored significantly higher than bands in classes 1, 2, and 3, and bands in class 4 scored significantly higher than bands in class 2 (see Table 7). The second post hoc analysis determined that orchestras in class 6 significantly outscored orchestras in all other classifications except class 3 (see Table 8).

Table 7. Mann-Whitney U Comparisons of Band Ratings by Classification

| Grade Level | N | U | p^a |
|-------------|-----|---------|-------|
| 1 vs. 2 | 149 | 1424.5 | .665 |
| 1 vs. 3 | 187 | 1690.5 | .244 |
| 1 vs. 4 | 169 | 1258.5 | .016 |
| 1 vs. 5 | 120 | 709.0 | .001 |
| 1 vs. 6 | 67 | 204.5 | <.001 |
| 2 vs. 3 | 288 | 9216.5 | .128 |
| 2 vs. 4 | 270 | 6782.5 | <.001 |
| 2 vs. 5 | 221 | 3783.5 | <.001 |
| 2 vs. 6 | 168 | 1043.0 | <.001 |
| 3 vs. 4 | 308 | 10071.5 | .014 |
| 3 vs. 5 | 259 | 5760.5 | <.001 |
| 3 vs. 6 | 206 | 1744.0 | <.001 |
| 4 vs. 5 | 241 | 6112.5 | .070 |
| 4 vs. 6 | 188 | 1950.0 | <.001 |
| 5 vs. 6 | 139 | 1541.5 | .003 |

^a Bonferroni correction for multiple comparisons ($\alpha = .003$)

Table 8. Mann-Whitney U Comparisons of Orchestra Ratings by Classification

| Grade Level | N | U | p^a |
|-------------|-----|--------|-------|
| 1 vs. 2 | 108 | 940.0 | .577 |
| 1 vs. 3 | 128 | 1079.0 | .226 |
| 1 vs. 4 | 107 | 929.5 | .568 |
| 1 vs. 5 | 82 | 653.0 | .615 |
| 1 vs. 6 | 60 | 262.0 | .001 |
| 2 vs. 3 | 188 | 3519.0 | .009 |
| 2 vs. 4 | 167 | 3028.0 | .099 |
| 2 vs. 5 | 142 | 2128.0 | .152 |
| 2 vs. 6 | 120 | 852.0 | <.001 |
| 3 vs. 4 | 187 | 4005.5 | .323 |
| 3 vs. 5 | 162 | 2786.0 | .346 |
| 3 vs. 6 | 140 | 1390.0 | .004 |
| 4 vs. 5 | 141 | 2395.5 | .956 |
| 4 vs. 6 | 119 | 992.0 | <.001 |
| 5 vs. 6 | 94 | 690.0 | .001 |

^a Bonferroni correction for multiple comparisons ($\alpha = .003$)

Discussion

This study compared the ratings and interrater reliability among multiple instrumental ensemble types, grade levels, and classifications using data from school band and orchestra festivals sponsored by the VBODA in 2010. The following discussion considers possible explanations for these findings and the implications they might hold for contests of the VBODA and similar organizations. Readers should generalize these results with caution, however, due to the differences in procedures and adjudication from one festival sponsor to another (e.g., Barnes & McCashin, 2005).

Distribution of Ratings

It is unclear why the average final rating (1.58) and frequency of Is (50.6%) and IIs (40.9%) was so high on a scale that included five potential ratings. The VBODA and other organizations (e.g., Boeckman, 2002; Hash, in press) should examine reasons for this phenomenon and decide if current ratings accurately assess differences in quality between participating ensembles. If not, festival organizers might experiment with different adjudication systems or raise the standard for I and II ratings. They might also consider dropping the IV and V ratings, which would place a II in a more appropriate context (e.g., Illinois Grade School Music Association Northern Division, 2010) without lowering final ratings for individual groups.

Although judges in this study may have consciously or subconsciously awarded high scores to encourage students and directors (e.g., Boeckman, 2002), it is more likely that these ratings reflected the quality of the ensembles and that only the best bands and orchestras in the state attended the festivals. If so, the VBODA might reach out to teachers who do not enter their groups and encourage them to perform for comments only. This effort would benefit the students and directors of these ensembles and, perhaps, inspire conference sessions or other programs (e.g., Indiana Bandmasters Association, 2011; Presley, n.d.) to assist instrumental teachers in challenging situations. Of course, a number of capable groups probably declined participation for reasons other than ability, such as the philosophical position of the director or administration regarding competition or the expense involved in entering and attending the event. Regardless, only 3.1% ($n = 31$) of groups ($N = 985$) entered VBODA festivals for comments only in 2010.

As with previous research (Orman et al, 2007; Yarbrough, Orman, & Neill, 2007), sight-reading scores in this study also reflected low variability and high marks, with 97.5% of participating ensembles earning a I (77.8%, $n = 568$) or II (19.7%, $n = 144$) rating. This result may have been due, in part, to the fact that some bands and orchestras ($n = 255$) chose the option that does not require sight-reading, or participated in class 1 and did not receive a rating for this event. Nonetheless, VBODA officials might consider the possibility that the sight-reading material is too easy and does not differentiate ensembles of varying levels of ability in this area, or try adding a second adjudicator to compensate for measurement error. Based on data for interrater reliability, Hash (in press) determined that two adjudicators was an appropriate number for sight-reading in high school band contests in South Carolina.

Interrater Reliability

Interrater reliability data indicated a moderately high level of agreement among pairs of adjudicators ($IRA_{pw} = .76$). Although this measure did not achieve the minimum benchmark of .80 for good reliability, it is possible that sites with the lowest IRA_{pw} included several bands and orchestras whose performance was on the border between two ratings (e.g., a low I or a high II) and that adjudicators' grades were closer than they appeared. A more precise measurement of reliability in these situations would require an examination of individual adjudication forms or the use of a point system that adds up to a total score. Still, the VBODA might examine adjudication in these panels to determine if individual judges or combinations of judges created widespread disagreement among ratings. Ideally, all judges would demonstrate a high level of reliability both individually and in combination with others. Forbes (1994), however, stated that identifying reliable evaluators was difficult because a standardized procedure for certifying adjudicators had yet to be developed.

Other reliability measures were much higher than IRA_{pw} . IRA_{co} reached at least .80 at all sites and .87 overall, suggesting that three-member judging panels provided a system of checks-and-balances that compensated for measurement error and helped insure a fair result. Additional analysis confirmed that reliability remained high, regardless of age level, ensemble type, or classification. Furthermore, internal consistency attained an alpha of .87, meaning that individual adjudicators were reliable in how they determined final ratings.

The fact that IRA_{pw} and IRA_{co} were so highly correlated ($r = .90$) indicates that these readings were probably redundant and that one measure of IRA, together with Cronbach's alpha, are probably sufficient for measuring interrater reliability in future analyses. IRA_{co} is probably the best indicator of IRA because this coefficient accounts for the checks-and-balances created by three-member adjudication panels. Together, alpha and IRA_{co} will account for the reliability of individual judges as well as adjudication panels as a whole.

Group Differences

The data does not indicate why middle school bands scored significantly lower than all other groups or why ensembles that played the most challenging literature generally outscored those in lower classifications, especially among bands. Although this result may be due to adjudicator bias against smaller or less experienced ensembles (e.g., Sullivan, 2003), high reliability for all contest sites, age levels, ensemble types, and classifications ($IRA_{co} > .80$), as well as relatively low correlations between classification and final ratings (bands, $r_s = -.32$; orchestras, $r_s = -.16$), suggests that ratings were probably accurate and reflected musical quality. Baker (2004) suggested that directors who choose more difficult repertoire might also set higher standards for performance and that ratings could be an indication of a director's expectations for the group. Nonetheless, these findings concur with previous research (Baker, 2004; Brakel, 2006; Hash, in press) that found significantly higher ratings for groups performing selections at a higher level of difficulty.

Analysis also indicated significantly higher ratings for sight-reading vs. concert-performance among the ensembles that participated in both events for a rating ($N = 730$). Perhaps groups who chose not to sight-read were among the least proficient in this area or judges applied a more lenient standard to sight-reading compared to the concert-performance segment of the festival. It is also possible that several ensembles were on the border between two ratings (e.g., low I or high II) and that the single adjudicator for this event awarded the higher rating to avoid grading too severely. Using two judges for sight-reading might help prevent this phenomenon and more accurately measure groups whose performance does not clearly indicate a specific rating.

Data from this study indicated a high level of interrater reliability for concert festivals sponsored by the VBODA, regardless of contest site, ensemble type, age level, or classification. Although final ratings differed significantly by age (middle school vs. high school band), ensemble type (middle school band vs. orchestra), and classification (lower vs. higher), these findings were probably the result of performance quality rather than adjudicator bias, since reading for interrater reliability remained consistent regardless of these variables. All organizations that sponsor large-group festivals should examine the statistical reliability of these events in order to insure that final ratings are fair and consistent. Future analyses also should focus on the correlation of grades given in individual categories and final ratings (e.g., Garman, Boyle, & DeCarbo, 1991), the effect of various evaluation forms on reliability (e.g., Norris & Borst, 2007), and the quality and consistency of written comments (e.g., Elis, 2007). Additional studies of directors, students (e.g., Austin, 1988; Stamer, 2004, 2006), and parents are needed to measure their attitudes toward festivals and determine why some teachers choose not to enter these events. Philosophical inquiry, in addition, might establish the extent to which contests are supporting instrumental programs in light of current education reform (e.g., Learning Point Associates, 2010) and changing attitudes towards the arts (e.g., Williams, 2011). Further research also should examine the effectiveness of current rating and adjudication systems to determine if revisions in contest procedures are necessary. Music organizations and researchers should work together throughout this process to insure that school music festivals serve teachers and students as much as possible.

References

- Adler, S. A., & Clark, R. (2008). *How it's done: An invitation to social research* (3rd ed.). Belmont, CA: Thompson Wadsworth.
- Austin, J. R. (1988). The effect of music contest format on self-concept, motivation, achievement, and attitude of elementary band students. *Journal of Research in Music Education*, 36, 95-107.
- Baker, V. (2004). The effect of repertoire selection on university interscholastic league choral concert ratings. *Texas Music Education Research*. Retrieved from <http://www.tmea.org/assets/pdf/research/Bak2004.pdf>
- Barnes, G. V., & McCashin, R. (2005). Practices and procedures in state adjudicated orchestra festivals. Update: *Applications of Research in Music Education*, 23(2), 34-41. doi:10.1177/787551233050230020105
- Batey, A. L. (2002, December). Take the terror out of adjudication. *Teaching Music*, 10, 40-46.
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51, 137-150. doi:10.2307/3345847
- Bergee, M. J. (2007). Performer, rater, occasion, and sequence as sources of variability in music performance assessment. *Journal of Research in Music Education*, 55, 344-358. doi: 10.1177/0022429408317515
- Bergee, M. J., & McWhirter, J. L. (2005). Selected influences on solo and small-ensemble festival ratings: Replication and extension. *Journal of Research in Music Education*, 53, 177-190. doi: 10.1177/002242940505300207
- 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, 34 -353. doi: 10.2307/3345660
- 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, 358-374. doi: 10.1177/002242940505300407
- Boeckman, J. (2002). Grade inflation in band contest ratings: A trend study. *Journal of Band Research*, 38(1), 25-36.
- Boyle, J. D. (1992). Program evaluation for secondary school music programs. *NASSP Bulletin*, 76(544), 63-68.

- Brakel, T. D. (2006). Inter-judge reliability of the Indiana State School Music Association high school instrumental festival. *Journal of Band Research*, 42(1), 59-69.
- Burnsed, V., Hinkle, D., & King, S. (1985). Performance evaluation reliability at selected concert festivals. *Journal of Band Research*, 21(1), 22-29.
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage, 1979.
- Cassidy, J. W., & Sims, W. L. (1991). Effects of special education labels on peers' and adults' evaluations of a handicapped youth choir. *Journal of Research in Music Education*, 39, 23-34. doi: 10.2307/3344606
- Cavitt, M. E. (1997). Effects of expectations on evaluators' judgments of music performance. *Texas Music Education Research*. Retrieved from http://www.tmea.org/080_College/Research/Cav1997.pdf
- Cavitt, M. E. (2002). Differential expectation effects as factors in evaluations and feedback of musical performance. *Texas Music Education Research*. Retrieved from http://www.tmea.org/080_College/Research/Cav2002.pdf
- Ciorba, C. R., and Smith, N. Y. (2009). Juries using a multidimensional assessment rubric measurement of instrumental and vocal undergraduate performance juries. *Journal of Research in Music Education*, 57, 5-15. doi: 10.1177/0022429409333405
- Conrad, D. (2003). Judging the judges: Improving rater reliability at music contests. *NFHS Music Association Journal*, 20(2), 27-31.
- Elliott, C. A. (1995/1996). Race and gender as factors in judgments of musical performance. *Bulletin of the Council for Research in Music Education*, 127, 50-55.
- Ellis, M. C. (2007). An analysis of taped comments from a high school jazz band festival. *Contributions to Music Education*, 34, 35-49.
- Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education*, 23, 186-189. doi: 10.2307/3344643
- Fiske, H. E. (1983). Judging musical performance: Method or madness? *Update: Applications of Research in Music Education*, 1(3), 7-10.
- Flores, R. G., & Ginsburgh, V. A. (1996). The Queen Elisabeth musical competition: How fair is the final ranking? *The Statistician*, 45(1), 97-104.

Hash

- Forbes, G. W. (1994). Evaluative music festivals and contests-are they fair?. *Update: Applications of Research in Music Education*, 12(2), 16-20. doi: 10.1177/875512339401200203
- Freelon, D. (2009). ReCal: Reliability calculation for the masses [online calculator]. Retrieved from <http://dfreelon.org/utills/recalfront/>
- Garman, B. R., Boyle, J. D., & DeCarbo, N. J. (1991). Orchestra festival evaluations: Interjudge agreement and relationships between performance categories and final ratings. *Research Perspectives in Music Education*, 19-24.
- Hash, P. M. (in press). Interrater reliability of high school band contest ratings in South Carolina. *Journal of Research in Music Education*.
- Hunter, D., & Russ, M. (1996). Peer assessment in performance studies. *British Journal of Music Education*, 13, 67-78.
- Illinois Grade School Music Association Northern Division (2010). *Constitution & bylaws, adjudicator rules, comment sheet samples*. Retrieved from <http://www.igsma.org/PDF/Addenda%20-%20Constitution/2010%20Revised%20Constitution-Bylaws.pdf>
- Indiana Band Masters Association (2011). Indiana band director mentoring project. Retrieved from http://www.indianabandmasters.org/Organization/mentor_project.htm
- Killian J. N. (1998). Characteristics of successful choirs in a contest setting. *Texas Music Education Research*, 39-43. Retrieved February 18, 2010, from http://www.tmea.org/080_College/Research/Kil1998.pdf
- Killian J. (1999). Music selection of successful choirs at UIL and non-UIL contests. *Texas Music Education Research*, 51-56. Retrieved February 18, 2010, from http://www.tmea.org/080_College/Research/Kil1999.pdf
- Killian, J. (2000). Effect of music selection on contest ratings: Year three of a continuing study. *Texas Music Education Research*. Retrieved February 21, 2010, from http://www.tmea.org/080_College/Research/Kil2000.pdf
- King, S. E., & Burnsed, V. (2009). A study of the reliability of adjudicator ratings at the 2005 Virginia band and orchestra directors association state marching band festivals. *Journal of Band Research*, 45(1), 27-32.
- Kinney, D. W. (2009). Internal consistency of performance evaluations as a function of music expertise and excerpt familiarity. *Journal of Research in Music Education*, 56, 322-337. doi:10.1177/0022429408328934

- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Latimer, M. E., Bergee, M. J., & Cohen, M. L. (2010). Performance assessment rubric reliability and perceived pedagogical utility of a weighted music performance assessment rubric. *Journal of Research in Music Education*, 58, 168-183. doi: 10.1177/0022429410369836
- Learning Point Associates (2010). *Evaluating teacher effectiveness: Emerging trends reflected in the state phase 1 race to the top applications* (Naperville, IL: Author).
- McPherson, G. E., & Thompson, W. F. (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education*, 10, 12-24. doi: 10.1177/1321103X9801000102
- Morrison, S. J., Price, H. E., Geiger, C. G., & Cornacchio, R. A. (2009). The effect of conductor expressivity on ensemble performance evaluation. *Journal of Research in Music Education*, 57, 37-49. doi:10.1177/0022429409332679
- National Music Adjudication Coalition (n.d.). Concert band or orchestra – music assessment form. Retrieved from <http://www.menc.org/documents/nmac/NMACbandorchestraform.pdf>
- Norris, C. E., & Borst, J. D. (2007). An examination of the reliabilities of two choral festival adjudication forms. *Journal of Research in Music Education*, 55, 237-251. doi:10.1177/002242940705500305
- Orman, E. K., Yarbrough, C., Neill, S., & Whitaker, J. A. (2007). Time usage of middle and high school band directors in sight-reading adjudication. *Update: Applications of Research in Music Education*, 25(2), 36-46. doi: 10.1177/8755123307250020105
- Phillips, K. H. (2008). *Exploring research in music education and music therapy*. New York: Oxford University Press.
- Presley, D. (n.d.). *The mentor project*. Retrieved from <http://www.scmea.net/mentoring.htm>.
- Price, H. E., & Chang, E. C. (2005). Conductor and ensemble performance expressivity and state festival ratings. *Journal of Research in Music Education*, 53, 66-77. doi:10.1177/002242940505300106
- Rickles, D. A. (2009). *A multivariate analysis of nonperformance variables as predictors of marching band contest results*. (Doctoral dissertation). Available from Dissertations and Theses database. (UMI No. 3353883)

Hash

- Rohrer, T. P. (2002). The debate on competition in music in the twentieth century. *Update: Applications of Research in Music Education*, 21(1). doi:10.1177/87551233020210010501
- Silvey, B. A. (2009). The effects of band labels on evaluators' judgments of musical performance. *Update: Applications of Research in Music Education*, 28(1), 47-52. doi:10.1177/8755123309344111
- Stamer, R. A. (2004). Choral student perceptions of the music contest experience. *Update: Application of Research in Music Education*, 22(2), 5-12. doi:
- Stamer, R. A. (2006). Changes in choral student perceptions of the music contest experience. *Update: Applications of Research in Music Education*, 25(1), 46-55. doi: 10.1177/87551233060250010106
- Sullivan, T. M. (2003). Factors influencing participation of Arizona high school marching bands in regional and state festivals. (Doctorial dissertation). Available from Dissertations and Theses database. (UMI No. 3080892)
- VanWeelden, K., & McGee, I. R. (2007). The influence of music style and conductor race on perceptions of ensemble and conductor performance. *International Journal of Music Education*, 25, 7-17. doi: 10.1177/0255761407074886
- Virginia Band and Orchestra Directors Association (2010). *Handbook*. Fairfax, VA: Author.
- 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, 34-39.
- Williams, D. A. (2011). The elephant in the room. *Music Educators Journal*, 98(1), 51-57.
- Yarbrough, C., Orman, E. K., & Neill, S. (2007). Time usage by choral directors prior to sight-singing adjudication. *Update: Applications of Research in Music Education*, 25(1), 27-35.

CONTRIBUTORS

PHILIP M. HASH serves as Associate Professor of Music Education at Calvin College in Grand Rapids, Michigan. He holds an Ed.D. in music education from the University of Illinois, a M.M. in wind conducting from Northwestern University, and a B.M. in music education from Millikin University. Recent publications include articles in the *Journal of Historical Research in Music Education*, the *Journal of Research in Music Education*, and the *Journal of Band Research*. Phillip taught instrumental music in northern Illinois for 14 years and has published instrumental arrangements with Grand Mesa Music, BRS Publications, and Daehn Publications.

After receiving his Bachelor's degree from Northwestern State University in 1990, **JEFFREY C. MATHEWS** taught high school band for seven years in the public schools of Louisiana and Texas. Dr. Mathews received the Master of Music Education degree from the University of North Texas, where he served as a Teaching Fellow. He continued his studies with Thomas Frascillo at the University of Southern Mississippi where he earned his Doctor of Musical Arts Degree in conducting pedagogy. He has been on the faculty of Northwestern State University since 1998 where he is an Associate Director of Bands and an Associate Professor of Music. He has served as Assistant Director of Bands, Director of Athletic Bands, Director of Student Activities and Organizations, and Associate Director of Bands while at Northwestern State. He currently teaches Band, Marching Band Techniques, Music Theory, Conducting, and Music Education. His research has published research articles and he has presented clinics at the Louisiana Music Educators Association Convention and the Texas Music Educators Association Convention. Dr. Mathews served for 15 years as the Commander and Conductor of the Texas Air National Guard Band. He led this band on a tour of the Czech Republic during the summer of 1999 and to Afghanistan and Iraq in 2007. He has recently been appointed Chief, Air National Guard Bands. In this capacity he advises, directs policy, and provides budgets for all the Air National Guard Bands in the United States.

GREG HANDEL joined the faculty in the School of Creative and Performing Arts at Northwestern State University of Louisiana in the Fall of 2008 as Assistant Professor of Music Education. Prior to his appointment at Northwestern State, Handel held the position of Instructor of Music and Assistant Director of Bands at Augustana College (SD), and served Arizona State University as a Graduate Teaching Assistant in the Sun Devil Bands program, as well as serving as a Graduate Intern in the Music Education Practicum class at ASU. He has taught all levels of instrumental music in both the public schools of Sioux Falls, SD, and Tempe, AZ. He also served the Tempe Elementary School District as instrumental representative to the Music Advisory Committee, curriculum coordinator for beginning band, and as a mentor teacher. He was also organist and choral ensembles accompanist at King of Glory Lutheran Church in Tempe, AZ, for seven years.

Dr. Handel maintains a special interest in historical research in music education. He has been published in the *Journal of Historical Research in Music Education*, and the *Music Educator's Journal*, and recently collaborated on a project presented at the International Symposium for Assessment in Music Education held in Bremen, Germany, and presented at the Chattanooga Symposium on the History of Music Education. He was awarded the 'Excellence In Teaching