



142
317
THS



This is to certify that the
thesis entitled

The Effect of Primary Performing Instruments
on Peer Evaluation

presented by

Bradford P Howells

has been accepted towards fulfillment
of the requirements for the

Master of Music degree in Music Education

Cynthia Crump Taggart
Major Professor's Signature

10/26/09

Date

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

The Effect of Primary Performing Instruments on Peer Evaluation

by

Bradford P Howells

A THESIS

**Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of**

MASTER OF MUSIC

Music Education

2009

ABSTRACT

THE EFFECT OF PRIMARY PERFORMING INSTRUMENTS ON PEER EVALUATION

by

Bradford P Howells

The purpose of this study was to investigate the validity of peer evaluation of solo performances of high school band students. The findings may be useful to a band teacher to enhance students' musical development, and ultimately their performance achievement. The problems of this study were: 1) to determine if high school students' peer evaluations of solo performances were valid when using a standard testing tool and 2) to determine if the student evaluator validity was different when the evaluator played the instrument being rated than when the evaluator did not play that instrument. The subjects in this study were high school band students (n=59) from a low-to-middle class, urban school district. Each student observed seven video-recordings of peer solo instrumental performances. Some of these performances were on the same instrument that the evaluating student played in band. Three expert musicians evaluated all the solo performances. There was a low to moderate correlation between student and expert evaluations and there was a significant difference found between same-instrument and not-same instrument classifications.

ACKNOWLEDGEMENTS

The completion of this thesis was greatly assisted by the efforts of the following people:

My wife, Amber, and my children who most of all sacrificed time and finances to help me complete this project.

My colleague and friend, Shawn Gurk, who listened to the first brainstormings of this project and helped at all levels in the development of the study.

My research advisor, Dr. Cynthia Taggart, who encouraged me to pursue a thesis track, invested hours of reading and analysis support.

My research committee, Dr. John Kratus and Dr. Gordon Sly, who reviewed my work and held me accountable to the standards of the profession.

My colleagues William Bier, Sharon Claassen, Jennifer Culler, and Aaron Good, Laura Hyler, and Lynn Potter who spent extra time evaluating solo performances.

The students of Wyoming Park High School Band who gave their time and energy in completing the evaluations.

The students of MSBOA District 10 who gave me permission to video tape and use their performances in this study.

Words and acknowledgements cannot convey the deep gratitude I feel for each of these individuals and groups. Your dedication to music and education is what fuels this project. Thank you. I hope my words do justice to the work you do.

TABLE OF CONTENTS

LIST OF TABLES	v
CHAPTER	
1. LITERATURE REVIEW	1
The Need for Evaluation.....	1
The Effects of Evaluation	3
Student Evaluations	4
Factors in Evaluation	7
Purpose and Problems.....	10
2. RELATED RESEARCH	11
Student Accuracy	11
Instrument Influence	16
Evaluator Experience	17
Summary	18
3. METHOD	20
Subjects	20
Design	20
Materials	21
Measures	24
Procedures.....	25
4. RESULTS AND INTERPRETATIONS	27
Means, Standard Deviations, and Correlation Factors.....	27
Discussion	28
5. CONCLUSIONS AND RECOMMENDATIONS	30
Purpose.....	30
Problems	30
Summary	30
Implications for Practice	31
Suggestions for Future Research	33
APPENDIX	
A. Woodwind Brass Solo Evaluation Form.....	35
REFERENCES	37

LIST OF TABLES

TABLE 1 - <i>n</i> of Recordings.....	22
TABLE 2 - Solo Performances per Evaluation Session	23
TABLE 3 – Means, Standard Deviations, and Correlation Factor	28

Chapter One – Literature Review

The Need for Evaluations

Music teachers have long used informal assessments, such as observation, to make mental note of how students are progressing at acquiring musical skills. However, large ensembles, limited class time, and performance pressures often force teachers to forget the specific attention to the evaluation of student learning that each individual student needs. As a result, the teacher may neglect to assess students' cognitive learning in a formal way. When the semester finishes, such a teacher will be forced to assign grades based on the last few observations and interactions with a student, which may or may not accurately reflect the student's true performance in class.

Occasionally, a participation grade will be weighted more heavily to boost a poor academic skill grade. McCoy (1991) found that band directors placed an average of 56.48% of their grade on attitude (affective) and concert participation and behavior (non-music). The choral directors in the same study based 55.67% of their students' grades on the same categories, while their principals would have based the majority (57.23%) of the grade on musical knowledge (cognitive) and performance ability (psychomotor) (McCoy, 1991). This complacent attitude of ensemble teachers towards assessment has been accepted in school music programs for decades.

With the advent of No Child Left Behind (US Department of Education, 2002), schools are required to display adequate yearly progress through state assessments. School administrations seek to measure the growth of their students throughout the year with formal assessments to ensure that they are meeting their own goals and benchmarks towards maintaining and/or improving their current educational success. Ultimately,

teachers are held accountable for the learning of their students. While the implications of this accountability has been debated in many staff meetings and criticized behind closed staff lunchroom doors, the importance of improving our schools is undeniable. In many school systems, this drive to improve filters down into the arts programs and creates a need to measure student achievement and progress.

While using evaluation techniques is not new to the music education profession, music teachers are often not equipped with strategies for assessing their students' performances. Music teachers' undergraduate training commonly does not spend a significant amount of time on creation and implementation of a quality assessment program (Lillis, 2000). Rubrics, continuous scales, additive scales and other such assessment terminology are unfamiliar to many music teachers. Professional development conferences and workshops have only recently started to incorporate assessment as topic that is worthy of study for music teachers. In the State of Michigan, the association that organizes vocal festivals addressed this shift by replacing their assessment system in 2008 with rubric-style evaluation (Stegman, 2009). The need for better comprehension and application of performance assessment has encouraged a large outgrowth of studies investigating the effects and determining the best methods of evaluation (Bergee, 1993 & 1995; Fiske, 1975 & 1977; Hewitt, 2001, 2002, 2005, 2007; Hewitt & Smith, 2004; Morrison, Montemayor & Wiltshire, 2004; Saunders & Holahan, 1997).

Do students actually learn better or learn more if their performances are being evaluated on a regular basis? How often should students have their performance evaluated? Is it the case that, to ensure student learning, the school concerts in December

and May are no longer sufficient, nor the district and state concert festivals held in the spring? To begin answering questions like these, Bergee (1997) studied the acceptability of student peer-evaluations. The accuracy of student self evaluations on solo performances was investigated twice by Hewitt (2002; 2005). Morrison, Montemayor, and Wiltshire (2004) studied the effect of self evaluation on the students' attitudes towards music performance. In a similar vein, Hewitt (2001) investigated the effect of self evaluation on their attitude towards practicing. Can they improve their practicing after being evaluated?

The Effects of Evaluation

Hewitt (2001) studied whether modeling, self-listening, and self-evaluation had an effect on junior high instrumental music students' performance and attitude about practice. His findings suggest that, when students only evaluate their own solo performances, there was no significant improvement in performance. However, self-evaluation combined with listening to a recorded model did tend to improve performances. He also found that students' attitudes towards practicing were not affected positively or negatively by self-evaluating.

Likewise, Morrison, Montemayor, & Wiltshire, (2004) revealed that a recorded model had a positive impact on self-evaluation. This study suggested that listening to a recorded model also improved the performances of the modeled song. Interestingly, other unmodeled songs performed by this group also improved when compared to performances by ensembles that had no modeling experiences. This finding implies that perhaps the effects of modeling transferred across performances.

Students in this study demonstrated increased discrimination of errors in their own performances in addition to increased awareness of expression and phrasing after having listened to a recorded model in the course of learning their pieces. The authors stated the benefits of this study as following: “Developing habits of self-evaluation in students is generally seen as desirable among music teachers as a means of encouraging student responsibility for musical learning” (Morrison, Montemayor, & Wiltshire, 2004, p. 118).

Student Evaluations

From the school administrators’ perspective, evaluations and assessments are the responsibility of the teacher and should objectively measure the growth demonstrated by the students. In this light, music teachers need to perform regular assessments to monitor growth and adapt teaching strategies for proper instruction. However, evaluation can be used as a curricular tool when students learn to evaluate performance, as mentioned previously (Wells, 1997). The National Music Education Standards, as created by MENC: the National Association for Music Education, include “Evaluating Music and Music Performances” as the seventh of nine standards. The description of this standard for students in grades five through eight says that students should:

- a) develop criteria for evaluating the quality and effectiveness of music performances and compositions and apply the criteria in their personal listening and performing.
- b) evaluate the quality and effectiveness of their own and others' performances, compositions, arrangements, and improvisations by

applying specific criteria appropriate for the style of the music and offer constructive suggestions for improvement (Music Educators National Conference, 1994, <http://www.menc.org/publication/books/prek12st.html>).

The description for grades nine through 12 is worded only slightly differently. Teachers as a whole agree that evaluation of musical performances is a skill that music students must have.

Investigators have explored how accurate students' evaluations are when compared to professional educators or adjudicators (Bergee, 1993, 1997; Byo & Brooks, 1994; Hewitt, 2005), as well as how students' evaluations change over time (Aitchinson, 1995; Hewitt, 2002). Establishing the accuracy of student evaluations is critical at the onset of any student-focused assessment system. If students' evaluations are not accurate, instruction on how to evaluate performance is necessary. The information gathered from such evaluations will have negligible educational value if they lack accuracy. Hewitt (2002) found that middle school students did not increase their ability to evaluate when using a self-guided evaluation form. Aitchinson (1995), however, found that, with teacher support, students did improve in their ability to self-evaluate. If accuracy can be improved over time and with guidance and instruction, there may be a number of uses for incorporating student evaluations into the music program.

First, students could develop the skill to listen critically to performances. This skill is one that all musicians desire for improving their own performances. Critical listening also enables musicians to gain experience from the performances of others. Second, students might be able to learn from the positive musical performances as well as

mistakes that they hear and, subsequently, improve their own performance achievement. Finally, students might develop the skill of communicating to others about their observations in a helpful manner.

Several studies have investigated student accuracy in a variety of settings. Hewitt suggested that middle school students tend to overrate their own solo performances when compared to expert raters (2002). High school students were only slightly more accurate than middle school students in certain sub-areas of evaluation, such as tone, intonation, tempo, interpretation, technique/articulation (2005). His 2002 study also showed that students participating in the study increased in their performance scores, but not in their ability to self-evaluate. After six weeks, some post-test sub area correlations improved slightly from the pre-test, which suggests that, over a longer time period and with more experience, self-evaluation accuracy might improve. The author proposes “that extended and perhaps more frequent opportunities should be offered for self-evaluation” (Hewitt, 2002).

Another study of junior high students’ abilities to evaluate full ensemble performance supported the claim that students had low correlation when compared to the rating of music educators ($r = .18$) (Byo & Brooks, 1994). However, this study also showed student ratings of a university level ensemble were more moderately correlated ($r = .50$) with ratings of experts. Temporal graphs presented similar ratings across time (Byo & Brooks, 1994). The authors suggested that perhaps the ability level of the performing ensemble affects the evaluation skills of students. This suggestion was supported by the weak correlation of student to expert evaluations, possibly because students were less objective in their evaluations of their own performances or because

they did not have the requisite musical skills to accurately evaluate their own performance (Byo & Brooks, 1994).

Two studies in a series conducted by Bergee (1993, 1997) on self-, peer-, and faculty evaluations of college-level solo performance corroborated previous findings. Correlations of self-evaluations with faculty evaluations were moderately low ($r = .10-.39$) in the 1993 study and moderate to inversely moderate ($r = -.54 - .56$) in the 1997 study. The peer-to-faculty evaluations resulted in considerably stronger correlations. In the 1993 study the correlations were ($r = .86-.91$) and the 1997 study ($r = .61-.98$) (Bergee, 1993, 1997).

These studies reveal that the self-evaluations of school age students should not be the sole method of evaluation, as those evaluations do not reliably correspond to those of music educators. Students may not be as objective as would be deemed ideal for the sake of assessment or they may not know enough to make valid performance assessments. Self-evaluation is a useful tool for developing critical listening skills, but a teacher should not assign a grade based upon student self-evaluations. Alternatively, peer-evaluations should be investigated further, as they may be more accurate and may be a practical tool in music education assessment.

Factors in Evaluation

When considering an individual who will be evaluating a performance, one must understand what personal characteristics may influence the evaluation. At solo and ensemble festival in the state of Michigan, the organizers of each event hire professional musicians to evaluate performances on their primary instrument. In select cases, a judge

may be asked to evaluate an instrument that is not his primary instrument, although in most occurrences, the instrument will be related to the primary instrument of the judge, such as clarinet to saxophone. It is assumed that the evaluation will be more accurate if done by someone who performs on the instrument and who has personal experience with its performance characteristics. Research shows that this may be an incorrect assumption (Fiske, 1975; Hewitt, 2007; Hewitt & Smith, 2004).

In a study comparing judges who performed on brass instruments with those who do not play brass instruments, there was no statistically significant difference between their ratings when evaluating high school solo trumpet performances (Fiske, 1975). Fiske also found that, when he re-categorized the judges as wind or non-wind instrument players, the only trait that resulted in a significant difference was technique. He suggested that, for purposes of auditioning for membership in an ensemble, the judges' primary performing medium need not be considered when selecting judges. However, the author did recommend that, in evaluations intended for improving a soloist's performance, it would be best to have a judge who at least played in the same type of performing ensemble, such as band or orchestra (Fiske, 1975).

Other studies supported this conclusion. When looking for significant relationships between experience level (lower-, upper-division college students, and in-service teachers) and primary performing instruments on evaluation reliability, Hewitt and Smith (2004) found few. The stronger relationships were between experience levels and not between primary performing instruments. The authors were led to concur with other studies that the performing instrument of the judge does not have any effect on the reliability of the evaluation (Hewitt & Smith, 2004).

In a similar study, Hewitt again investigated effects of age level and primary performing instruments on evaluation reliability. In this study, however, he grouped the students by middle school, high school, and college level. Again, he found no influence of primary performing instrument on the ratings at any age level. However, Hewitt did find some significant differences due to age level. One finding that was particularly interesting was that, overall, the middle and high school students rated the performances lower than did college students. These findings contrasted previous studies that suggested that younger students tend to overrate in evaluation settings (Byo & Brooks, 1994; Hewitt, 2002). The design of the evaluation may have had some influence in this, as students were not self-evaluating in this study but rather peer-evaluating (Hewitt, 2007).

As mentioned previously, professional musicians are hired to adjudicate at solo and ensembles. It is also assumed that evaluators must have a higher level of performance achievement than the performer to accurately evaluate student performances. One study shows that this also may be untrue. When looking for relationships between judge performance achievement, judge reliability, and judge non-performance achievement, Fiske found that there was no relationship between performance achievement and reliability of ratings or between performance achievement and non-performance achievement. Non-performance achievement was defined as the cumulative scores of the judges' college level music history and music theory classes. However, there was an inverse relationship between non-performance achievement and judge reliability. In other words, judges who do well in music history and music theory may actually be worse at evaluating performances. The author attributed this phenomenon to differing mental mechanisms used in various music disciplines.

“Disciplines that require absolute responses, such as music history and music theory, ordinarily would provide little practice for such a [discretionary] mechanism and, at worst, would tend to extinguish its use altogether. Conversely, teaching experience in performance would tend to strengthen the mechanism since student progress in performance depends upon ongoing evaluation” (Fiske, 1977).

Purpose and Problems

Limited research has been conducted on the accuracy of peer evaluations. There have been no studies that ask students to evaluate solo performances of multiple instruments. Therefore the purpose of this study is to investigate the validity of peer evaluation of solo performances of high school band students. The findings may be useful to a band teacher to enhance students’ musical development, and ultimately their performance achievement. The specific problems of this study are as follows:

1. Are high school students’ peer evaluations of solo performances similar to those of expert judges when using a standard testing tool?
2. Is the student evaluator accuracy related to whether the evaluator plays the instrument being rated?

Chapter Two - Related Research

Student Accuracy

Many studies have found that student self-evaluations of performances have little relationship to the evaluations of expert music educators (Bergee, 1993, 1997; Byo & Brooks, 1994; Hewitt, 2005). Bergee used similar methods in both of his studies to obtain the correlations of $r = .10-.39$ in a 1993 study and $r = -.54 - .56$ in a 1997 study between student self-ratings and the ratings of musical experts. The evaluated performances were of college-level students who were performing juries for a panel of faculty evaluators. The performances were video-recorded. The faculty evaluations were done in real-time, while the peer and self evaluations were completed while watching video recordings. Performances were already scheduled to occur, and the faculty members were well-acquainted with the evaluation process. To prepare materials for evaluation, the author merely needed to video-record the performances to allow for subsequent viewing by the performers. While this presented a practical solution for this study, differences in evaluations may have occurred due to the nature of the presentation of the performances (Bergee, 1993, 1997).

In the 1993 study, Bergee attempted to account for these discrepancies by using a technique involving comparisons of mean differences. The reported differences ranged from .03 to .51, which indicated relatively strong agreement (0 would indicate complete agreement of scores while 4 indicates complete disagreement) (Bergee, 1993).

When investigating self-evaluation accuracy, several researchers found that recording the performance first and then evaluating the recorded performance results in greater student objectivity than evaluating a completed live performance (Byo & Brooks,

1994; Hewitt, 2002, 2005). This may be due to the student evaluators' focus of attention during the performance. There may exist too great of a challenge on the part of the student to attend to both his/her current playing and also remember every aspect for later retrieval in the evaluation setting. Self-evaluating may also be viewed as assigning oneself a grade. Therefore, middle school students especially may not be able to objectively evaluate themselves if they see the evaluation in that light.

A recorded performance is also more practical when using a large number of evaluators (Bergee, 1997). The use of video recordings as opposed to audio recordings does not seem to affect the reliability of evaluations in any study. With the rapid development of high quality recording technology, video recording is no more complicated than audio recording. As a result, the resulting recorded performance may feel more authentic or "live" when video-recorded. For the purpose of this study, I chose to have both expert and student peer evaluations conducted using identical video-recorded materials. This choice allowed the evaluation format to be the same for both the students and the expert evaluators and contributed to the validity of the study.

The method of measurement has varied throughout the studies. Saunders and Holahan (1997) created criteria specific ratings scales and determine the accuracy their measures. They also studied whether these scales helped the judges differentiate between levels of performance. The results of this study revealed that the Woodwind Brass Solo Evaluation Form [WBSEF] has high internal reliability (.92) and has been shown to be effective when used by middle and high school students (Hewitt, 2001, 2002, 2005, 2007; Hewitt & Smith, 2004). The authors also found that WBSEF also allowed the judges to

specifically focus on areas of accomplishment and address areas where the performer needed assistance (Saunders & Holahan, 1997).

Bergee used a measure he created called the Brass Performance Rating Scale [BPRS], which included 27 statements that were categorized into four factors: interpretation/musical effect, tone quality/intonation, technique, and rhythm/tempo. Each item was rated in Likert format with 5 points per item. Some of the statements were positive and therefore earned 1 point for a strongly disagree to 5 points for strongly agree. The negative statements were scored 5 points for strongly disagree to 1 point for strongly agree. He referenced his own prior studies for reliability. Total score reliability in those studies was strong ($r = .94-.98$), as was reliability among factors ($r = .89-.99$). None of the statements referred to specific brass characteristics, and thus it may possible to use this measure for any wind instrument. However, length of time required to read and make a judgment on 27 separate statements may be counterproductive if the BPRS was used with high school students (Bergee, 1993). As a result, this study will use the Woodwind Brass Solo Evaluation Form [WBSEF] (Saunders & Holahan, 1997), which also has high internal reliability (.92) and has been shown to be effective when used by middle and high school students (Hewitt, 2001, 2002, 2005, 2007; Hewitt & Smith, 2004).

The reliability investigation conducted by Byo and Brooks (1994) showed that students were less reliable compared to expert raters when they evaluated their own ensemble's performance ($r = .19$) than when they listened to a university ensemble playing a similar style piece ($r = .50$). Two factors of their methodology must be taken into consideration in light of the present study. First, the authors chose to use a

Continuous Response Digital Interface (CRDI) to collect data on the listeners' reactions. This device allows evaluators to turn a dial to rate the overall quality of the performance on a scale up to 100. CRDI does not reveal any data concerning the dimension of the musical performance to which the evaluators are responding. Although data is coded for time, and this coding can be aligned with the performance, evaluators are often responding several seconds later than the actual event they are evaluating. The resulting data is interesting and useful for comparison, but it does not inform the readers beyond the graphs and numbers. The authors admitted that it must be assumed that the students were actually evaluating the quality of the performance and not rating their preference for the performance. Therefore, this study will include ratings of the specific dimensions of musical characteristics of tone, intonation, technique/articulation, melodic accuracy, rhythmic accuracy, tempo, and interpretation as included in the WBSEF.

The second factor to be considered in the Byo and Brooks study (1994) is that students were not evaluating solo performances. Listening to ensembles takes on a different form due to the harmonic textures and various timbres that occur. It cannot be assumed that students are sufficiently experienced in ensemble evaluation to accurately perform such a task (Byo & Brooks, 1994).

Another study that compared student self-evaluations with expert evaluators in an ensemble setting found that high school students' scores had no significant correlations to experts' scores in any subarea ($r = -.12 - .21$) (Hewitt, 2005). Hewitt found a low to moderate correlation for middle school students ($r = .20-.38$). The student participants in this study were asked to evaluate themselves after they had just finished performing a selected ensemble piece in a summer music camp rehearsal setting (Hewitt, 2005). The

accuracy of such evaluations must be questioned due to the methods used. Students, especially those in middle school, may find it challenging to distinguish their performance in each of the different musical subareas of WBSEF, as they are performing their part of a full wind ensemble. For example, most customary arrangements for middle school bands do not often present a significant portion of melodic material to the low brass instruments. French horns are often asked to play rhythmically demanding parts; yet they may not be able to identify how their part supports the rest of the ensemble. This may make it difficult to rate interpretation or melodic accuracy. Solo evaluations are simpler, as there is only one performer to consider.

Certainly, an accompanist may play a role in the overall performance, but this role can be minimized with a valid measurement instrument. One study has shown that student evaluators are able to identify strongest and weakest aspects of solo performances regardless of accompaniment style (Brittin, 2002). Solos are complete in and of themselves, without needing context of an entire ensemble. The present study will continue to look at the evaluations of solo performances.

While studies have found consistently low correlations (Byo & Brooks, 1994; Hewitt, 2005), both of the Bergee studies revealed strong relationships between peer and faculty evaluations. In the 1993 study, the correlations ranged from .86-.91 and in the 1997 study, the range was .61-.98 (Bergee, 1993, 1997). The greater range in the 1997 study was attributed to the combined factors of small sample size and large variety of solo performance instruments and large variety of faculty performance instruments. Specifically, there were five vocal, three string, four brass, four woodwind, and three percussion faculty from one site evaluating the performances. Solo performances

consisted of seven vocalists, six string players, eight brass players, nine woodwind players, and seven percussionists. Interestingly, the instruments with the strongest faculty-peer correlations had the weaker faculty-self and peer-self correlations (e.g., Percussion, Site 3 Faculty-Peer $r = .98$, Faculty-Self $r = -.19$, Peer-Self $r = .06$). The opposite was also true; the strongest faculty-self and peer-self correlations had the weaker faculty-peer correlations, although the stronger faculty-self and peer-self correlations were negative and the faculty-peer correlations were statistically strong (e.g., Strings, Site 1 Faculty-Peer $r = .75$, Faculty-Self $r = -.48$, Self-Peer $r = -.59$) (Bergee, 1997). To increase the chances that the method of the study does not interfere with the data, the number of expert evaluators will be limited to two brass and two woodwind experts, while the number of student evaluators will be maximized. Because the second problem of this study is specifically looking at the effect of the evaluator playing the same instrument as the performance being evaluated, this study will seek a broad representation of primary instruments.

Instrument Influence

A significant number of studies have shown that the primary performing instrument of the evaluator does not influence the accuracy of the evaluation (Fiske 1975; Hewitt, 2007; Hewitt & Smith, 2004). These studies were done only on trumpet performances for which the evaluators were grouped as brass or non-brass performers. All three found that there were no significant differences between brass and non-brass evaluators, including overall evaluations and evaluations for traits or subareas (Fiske 1975; Hewitt, 2007; Hewitt & Smith, 2004). However, at the results of studies that were

limited to trumpet performances alone do not necessarily generalize to other instruments. Trumpet performance traits may be more recognizable by a broad number of musicians, especially brass musicians; thus similar standards for performances may already exist. Considering this, the present study is designed so that the student evaluators of all instruments will evaluate performances of all instruments.

Evaluator Experience

Two recent studies have investigated the effect of the evaluator's age or experience level on the accuracy of their evaluations (Hewitt, 2007; Hewitt & Smith, 2004). Hewitt and Smith divided college students into lower and upper classmen, and compared these two categories with a third, in-service teachers. This study found no statistically significant difference between the ratings of the three experience levels. The evaluators were listening to performances of junior high trumpet players, and the differences emerged over one performer in particular. Upper-classmen rated this performer higher in tone and intonation than lower-classmen and in-service teachers. Lower-classmen also scored the intonation of a different performer significantly higher than upper-classmen. The authors of this study concluded that, for the study as a whole, experience had little influence on the evaluations. To explain this, they state: "The lower- and upper-division college students in this study seem to have reached the level of sophistication that allowed for them to evaluate a diverse sample of junior high trumpet players in a manner similar to more experienced teachers" (Hewitt & Smith, 2004, p. 324).

In a study with a similar design, Hewitt (2007) investigated the influences of education level on evaluation. He compared middle school, high school, and college level students. The results suggested that these age groups evaluate performances differently, especially when focusing on sub-areas. Tone was mostly rated lower by middle and high school students than college students. Evaluations between middle school and high school students were the most similar for the majority of performances and across subareas. Evaluations by these groups of students were more often lower in ratings than college age students (Hewitt, 2007).

Many other studies have used evaluators at various education and experience levels. Both studies by Fiske involved expert evaluators (1975, 1977). Both studies by Bergee (1993, 1997), one study by Hewitt (2007), and the study by Hewitt and Smith (2004) involved college students. High school evaluators were the focus of two of Hewitt's studies (2005, 2007). Several authors have used junior high students in evaluations (Byo & Brooks, 1994; Hewitt 2002, 2005, 2007). The results of these studies suggest that conducting a study among high school students should yield consistent results between similarly aged student evaluators.

Summary

I have taken into consideration prior research methods and results in the development of the design and methods for the present study. The following list contains a summary of these considerations:

- 1) Both expert and student peer evaluations should be conducted using identical materials and evaluation formats.

- 2) This study will use the **Woodwind Brass Solo Evaluation Form [WBSEF]** (Saunders & Holahan, 1997), which has high internal reliability (.92) and has been shown to be effective when used by middle and high school students (Hewitt, 2001, 2002, 2005, 2007; Hewitt & Smith, 2004).
- 3) This study will include ratings of musical characteristics of tone, intonation, technique/articulation, melodic accuracy, rhythmic accuracy, tempo, and interpretation as included in the WBSEF.
- 4) The present study will look at the evaluations of solo performances as opposed to full ensembles or individual performances within an ensemble.

Chapter Three - Method

Subjects

Subjects in this study were from a west Michigan school district that has a fairly diverse population and has a medium sized high school band program of approximately 80 members. The band program uses standard instrumentation, including all of the major solo performing instruments (flute, clarinet, alto saxophone, trumpet, F horn, trombone, snare and mallet percussion) and other performing instruments as they are available (oboe, bassoon, bass clarinet, tenor saxophone, euphonium, and tuba).

Students in this program are familiar with solo and ensemble and full ensemble concert festivals, as they participate in them on an annual basis. This type of evaluation comprises the majority of their prior evaluation experience. As a result, their band teacher uses the basic terminology of musical sub-areas on the ratings forms for those events during class. Their instruction consists of a comprehensive music education through performance so that the terms tone, intonation, rhythm, melody, and interpretation are familiar and functional vocabulary.

Design

The design of this study is two-fold. To answer the first question, I used a cross sectional design to determine correlations between student (peer) and expert evaluations of performances. I answered the second question using a non-statistical comparison of the correlation between the students and the expert judges when the student evaluator played the same instrument as the performance being evaluated (hereafter labeled “same-

instrument”) and when the students did not play the instrument of the performance being evaluated.

Materials

In order to answer the questions of this study, I needed to have recordings of solo performances on a variety of instruments that both the students and expert judges could rate to yield the data for the study. As solo-ensemble festival is the most common venue for solo performance, and the performances at those festivals are knowingly performed for ratings, taping solo-ensemble festival performances seemed logical. Prior to a solo and ensemble festival, I contacted the band directors of the schools within the district that normally would participate in the festival. I provided the directors with a consent form to distribute to their students, and asked the directors to provide me with performance schedules of the students who consented to participate. Due to their school responsibilities, only two directors returned schedules of consenting students from which I could create a schedule to record the performances. To remedy this shortage of participants, I approached groups of performers on the day of the festival, requesting to video-tape their performance for use in this study. Prior to the performance, the soloist and his/her parents completed the consent form. All video recordings were made using the same digital video recorder and recorded directly to the hard drive of a laptop computer. The number of solo performances obtained can be found in Table 1.

To increase the likelihood of a random sample, the researcher obtained performances from students who attended a variety of schools and represented a variety of grade levels. This reduced the effect of each school’s program or experience level on

the student evaluations. I also spread the recordings of each instrument throughout the day so that the evaluations were not affected by the time of the performance.

Ideally, I would have preferred to gather four video recordings per major solo instrument. While I only planned to use two or three recordings for evaluation, gathering more recordings would have allowed me to discard any recording that may have lower recording quality or technical difficulties. I had planned to obtain a single high-quality solo recording of other solo instruments (oboe, bassoon, bass clarinet, tenor saxophone, euphonium, and tuba) that are not as widely studied in the band classrooms.

Table 1

n of Recordings

Solo Instrument	<i>n</i> recordings
Flute	3
Clarinet	2
Bassoon	1
Alto Sax	2
Tenor Sax	1
Trumpet	3
F Horn	1
Euphonium	1
Tuba	1
Total	15

After gathering the recordings, I created five video compilations, each of which included seven solo performances and time in between for the student evaluators to complete their WBSEF forms. One video compilation was watched per evaluation session. Two video compilations could fit on one digital video disc (DVD), therefore three DVDs were created.

The seven performances included in each video were chosen based on three factors. 1) All recorded performances of the primary performing instrument for that session would be included in the video compilation. 2) The remaining performances would represent a variety of instruments that students in that session would not be as familiar with. 3) Each recorded performance would be included equally across all seven performances, to the extent possible. In each session, there were one to three performances on the primary instrument of the student evaluators. These were intermixed with four to six other performances of other various instruments. Each student evaluator observed seven total performances. The total number of students who performed evaluations was 59.

Student sessions were divided based upon primary performing instruments and were grouped to keep the number of students per session relatively similar. I desired to be sensitive to the band director's need for rehearsal time and minimum distractions to the week; thus smaller groups were also grouped so that requiring extra days of evaluations could be avoided. Trumpet and horn players were grouped together as were all low brass voices (trombone, euphonium, and tuba). A table that contains each DVD's contents can be seen below.

Table 2

Solo Performances per Evaluation Session

	Flute Session	Clarinet Session	Saxophone Session
Performance #1	Trumpet #2	Clarinet #2	Tenor Sax #1
Performance #2	Flute #2	Tuba #1	Flute #2
Performance #3	Tenor Sax #1	Flute #1	Alto Sax #1
Performance #4	Clarinet #2	Clarinet #1	Trumpet #1
Performance #5	Flute #1	Trumpet #2	Alto Sax #2
Performance #6	Euphonium	Alto Sax #1	Clarinet #2
Performance #7	Flute #3	F Horn #1	Bassoon #1

Table 2, continued

	Trumpet/Horn Session	Low Brass Session
Performance #1	Alto Sax #2	Bassoon #1
Performance #2	Trumpet #3	F Horn #1
Performance #3	Clarinet #1	Tuba #1
Performance #4	Trumpet #1	Clarinet #1
Performance #5	Flute #3	Euphonium #1
Performance #6	Trumpet #2	Flute #1
Performance #7	F Horn #1	Alto Sax #2

I created a separate DVD that contained all 15 performances which were observed by the expert judges in random order.

Measures

Both the students and expert judges used Woodwind Brass Solo Evaluation Form [WBSEF] to rate the performances because of the strong reliability as a whole (.92) and across the range of instruments (.82-.97) as documented in previous research literature (Saunders & Holahan, 1997). In other studies as well, WBSEF has been shown to have a strong interjudge reliability (Hewitt, 2001, 2002, 2005, 2007; Hewitt & Smith, 2004). WBSEF has been used in many formal studies with middle and high school aged students and has been shown to be appropriate for use with performances of this age level.

To complete WBSEF, the evaluator is presented with criteria-specific, continuous five-point ratings scales in each of six sub-areas: tone, intonation, melodic accuracy, rhythmic accuracy, tempo, and interpretation. A seventh sub-area, technique/articulation, is rated using an additive five-point scale (See Appendix A). WBSEF also includes

rating scales for evaluating the performances of scales and sight-reading. In the context of this study, these were not used.

Procedures

Consent forms were given to the student participants from the band chosen to do the evaluations. The forms were taken home to be signed and returned. During their regularly scheduled class rehearsal time, I took one instrument group to a separate room to view the recordings. I distributed seven copies of the WBSEF to each student evaluator, who then watched a two-minute instructional video on how to use WBSEF that I created to keep the instruction as consistent as possible. It was noted in this video that the technique/articulation section of the form was additive, or “Check all that apply.” Some students expressed a concern about the wording of this section. The items read “as marked” for concepts such as accents, ornamentations, and articulations (see Appendix A). Because the students did not have the musical score, I advised them to check the selections if the accents, ornamentations, and articulations were played in a way that was musically appropriate. Finally, we discovered during the sessions that some of the video recordings had an audible “popping” sound that was a recording deficiency and not a property of the musical performance. I advised the students to disregard this in their evaluations. The data collected do not reflect any effect of this defect on the scoring.

According to the author of WBSEF, the evaluator is to act as a reporter and via the form and “describe the levels of performance achievement” (T.C. Saunders, personal communication, March 6, 2008). Evaluators were advised not to replace the numerical values for descriptor, such as excellent, good, average, or poor. I answered any questions

from **st**udents to ensure that optimal understanding was established prior to **comm**encement of the evaluation period.

The students evaluated each of the seven solo performances on separate forms in one **se**ssion. After evaluating the performances, student evaluators would fill out a short **surve**y to identify their primary performing instrument and any secondary instruments **the**y may perform on in other settings, such as marching band or jazz band.

After conducting the student peer-evaluation sessions, I contacted three **pro**fessional instrumental musicians to evaluate the recorded performances. Before the **eva**luation session, I informed the judges about the purpose and methods of this study. I **comp**ared the WBSEF to other evaluation tools these judges were familiar with to show **diff**erences during a short discussion about the use of the WSBEF. I pointed out similar **issu**es as discussed in the student sessions. Then, in one hour-long session, the judges **eva**luated all of the performances using WBSEF.

Chapter Four - Results and Interpretations

Means, Standard Deviations, and Correlation Factor

I calculated the inter-judge reliabilities between expert evaluators using a correlation matrix to determine how consistent the scores were between judges. The correlation between expert judge 1 and judge 2 was .78, between judges 2 and 3 was .75, and between judges 3 and 1 was .86. These correlations are within an acceptable range for inter-judge reliabilities.

After each of the instrument groups had evaluated the performances on their specific compilation, I analyzed the data to determine means and standard deviations of all evaluations, as well as the results according to the grouping of same instruments and not-same instruments. Some student evaluators indicated that they played multiple instruments in performance settings. For example, during the concert season, one student played trumpet. During marching season, however, this student played euphonium. In such cases, the student's primary instrument was classified by whichever session (flute, clarinet, saxophone, trumpet/horn, or low brass) he or she participated in during the evaluation process. The trumpet/euphonium player mentioned here was considered a trumpet player because she attended the trumpet/horn performance rating session. Data from the ratings of instruments that the student played during other times of the school year were excluded from the study.

The student evaluations were correlated with those of the expert judge scores for the same-instrument performances using the Pearson Product Moment formula. The resulting means, standard deviations, and correlations are reported in Table 3.

Table 3

Means, Standard Deviation, and Correlation Factor

	Student Mean	Student SD	Expert Mean	Expert SD	<i>r</i>
All Evaluations	55.48	9.48	55.57	6.25	.44
Same Instrument	55.87	9.29	54.82	6.95	.58
Not Same Instrument	55.23	9.59	54.15	5.58	.39

Discussion

The means of the students and the expert judges were similar, even when taking **in** to consideration same and not same instruments. However, the standard deviations of **the** students tended to be much larger than those of the expert judges. This means that **there** was more variance in the student scores than in those of the judges.

The first problem of this study was to determine if high school students' peer **e**valuations of solo performances agree with those of expert judges when using a standard **e**valuation instrument. The correlation between all student evaluations and expert **e**valuations is moderate to low ($r=.44$). This suggests that student evaluations may not **be** an accurate reflection of the quality of the performance. Although the moderate to low **c**orrelation found in this study is slightly higher than those of previous studies, the **P**ractical implications are much the same. Byo and Brooks (1994) found a low **c**orrelation ($r=.18$) when junior high students evaluated their own ensemble performance. **A**lthough Hewitt (2005) looked at each individual music performance subarea (tone, **i**ntonation, melodic accuracy, rhythmic accuracy, tempo, interpretation, and **t**echnique/articulation) the range of correlations he found ($r= -.12 - .21$) was somewhat **l**ower than that of this study but practically comparable.

The second problem of this study was to determine whether the student evaluator accuracy was affected by whether they were evaluating a performance of the primary performing instrument of the evaluator or of an instrument that the student did not play. The correlation with the ratings of expert judges to students who play the same instrument as the evaluated solo performance was moderate ($r=.58$), while the correlation between expert judges' ratings and the ratings of students listening to instruments that they did not play was considerably lower ($r=.39$). The resulting differences between the two correlations reveal that there is some effect of same-instrument versus not-same-instrument evaluation on students' abilities to rate performances. These findings are in moderate disagreement with previous studies, which suggested that the primary performing instrument had little effect or no affect on the overall solo evaluation (Fiske, 1975, Hewitt & Smith, 2004).

Chapter Five - Conclusions and Recommendations

Purpose

The purpose of this study was to investigate the validity of peer evaluation of solo performances of high school band students. The findings may be useful to a band teacher to enhance students' musical development, and ultimately their performance achievement.

Problems

The problems of this study were: 1) to determine if high school students' peer evaluations of solo performances were similar to those of expert judges when using a standard testing tool and 2) to determine if student evaluator accuracy is related to whether the evaluator plays the instrument being rated.

Summary

The importance for student musicians to be evaluated on their own performance, along with evaluating the performances of others, is becoming increasingly evident in the music education community. The National Standards reflect this by including "Evaluating Music and Music Performances" (Music Educators National Conference, 1994, <http://www.menc.org/publication/books/prek12st.html>). The focus of this study was on the validity of ratings when comparing evaluations of students who play the same instrument as the one being evaluated to those of when the student does not play that instrument.

The subjects in this study were high school band students (n=59) from a low-to-middle class, urban school district. Each student rated seven video-recordings of peer solo instrumental performances using WBSEF. Some of these performances were on the primary performing instrument of the evaluating student and others were not. Three expert musicians evaluated all the solo performances.

The correlations of the student-to-expert evaluations were calculated as a whole and then according to whether the student was evaluating a performance on the same instrument that he or she played in band as opposed to a different instrument. There was a moderate to low correlation between student and expert evaluations when the ratings were considered as a whole. However, the correlation between same instrument ratings and those of the expert judges were somewhat higher than those of the overall ratings or different instrument ratings.

Implications for Practice

The similarities of the moderate to low correlation factors amongst this and other studies (Byo & Brooks, 1994; Hewitt, 2005) investigating student evaluations suggests that most student musicians do not evaluate other solo performances very well. However, in this study, students were more accurate in their ratings when they were rating a performance on their primary instrument. The music education community places a high priority on students' abilities to evaluate performances, as mentioned previously. With teacher guidance, it may be possible to improve this ability (Aitchinson, 1995). Therefore, more classroom time should be spent guiding students in evaluation. Perhaps a bi-weekly performance evaluation session, in which three to five

students perform an exercise, a portion of the upcoming concert, or a recent Solo & **E**nsemble piece could be put into place to allow students a chance to evaluate their peers. **G**uided practice, teacher feedback, and class discussion might be helpful in teaching the **p**rocess.

It is encouraging to discover that instrument experience moderately affects peer **e**valuations. Logically, a student may appreciate some of the finer points of performing **o**n an instrument, especially in the area of technique, if a student has more experience **w**ith that instrument. During the evaluation process, this student may identify **p**erformance weaknesses that others would miss because he or she has encountered **s**imilar weaknesses in his or her own experience. Likewise, an expert musician may be **m**ore familiar with these tendencies. This may account for the slightly higher **c**orrelations for same-instrument ratings. Fiske's study (1975) corroborated this, when he **f**ound that only the area of technique showed a significant difference between evaluators **w**ho played the same type of instrument and those who did not.

The results of this study show that all students need guidance learning to transfer **t**he knowledge and experiences they acquire on their own instruments so that they can **a**pply them to performances on other instruments. They also need to continue to develop **t**heir evaluation skills on their own instruments. To help develop this, a teacher could **h**ave students perform for the class and have each student in the class evaluate the **p**erformance. The evaluating students gain valuable experience in assessing other **p**erformances, ideally improving their skills in evaluation with each attempt (Aitchinson, 1995).

Suggestions for Future Research

It is evident from this study and others in this field that music educators need to involve their students in the processes of evaluating musical performances. More research should be done to determine the extent to which students are able to evaluate musical performances and how these evaluations can be improved. Possible research topics could include peer solo evaluations that focus on specific musical dimensions such as tone, intonation, melodic accuracy, rhythmic accuracy, tempo, interpretation, and technique/articulation. Examining students' abilities to focus on these specific dimensions may reveal more about their musicianship and the areas upon which teachers need to focus when developing student evaluation skills. The current study gathered this kind of data, but it was not analyzed in such a way as to reveal the validity of the student ratings of each sub-area.

Also, do students evaluate full ensemble performances more accurately than solo performances? Considering that they are engaged in full ensemble performance more often than solo performance, this may be a revealing study. While solo performances are ideal for assessing individual growth in a student, a large number of student musicians may not seek the opportunity to perform on their own. The majority of their musical experience will be in the full ensemble setting. Therefore, their evaluation abilities when listening to or performing in an ensemble may differ from those used in solo performances. Is there a difference, and, if so, is this differentiation important enough to be addressed?

In addition, WBSEF includes rating scales for playing scales and sight-reading, neither of which were used in the context of this study. Can students effectively evaluate their performances these areas using the WBSEF?

Finally, which experiences help students improve their skills in music evaluation? Can providing regular opportunities for students to evaluate performances increase the accuracy of their ratings? Will their ability to evaluate performance with more accuracy result in the development of richer musical skills? Studies that seek the answers to these questions are vital to the continued success of student musicians and music education in our schools.

Appendix A - WOODWIND/BRASS SOLO EVALUATION FORM

Evaluator Number: _____

Sample Number: _____

Final Score: _____

TONE The performer's tone: *(Check ONE only)*

- 10 ____ is full, rich, and characteristic of the tone quality of the instrument in all ranges and registers.
- 8 ____ is of a characteristic tone quality in most ranges, but distorts occasionally in some passages.
- 6 ____ exhibits some flaws in production (i.e., a slightly thin or unfocused sound, somewhat forced, breath not always used efficiently, etc.).
- 4 ____ has several major flaws in basic production (i.e., consistently thin/unfocused sound, forced, breath not used efficiently).
- 2 ____ is not a tone quality characteristic of the instrument.

INTONATION The performer's intonation: *(Check ONE only)*

- 10 ____ is accurate throughout, in all ranges and registers.
- 8 ____ is accurate, but performer fails to adjust on isolated pitches, yet demonstrates minimal intonation difficulties.
- 6 ____ is mostly accurate, but includes out-of-tune notes. The performer does not adjust problem pitches to an acceptable standard to an acceptable standard of intonation.
- 4 ____ exhibits a basic sense of intonation, yet has significant problems, performer makes no apparent attempt at adjustment of problem pitches.
- 2 ____ is not accurate. The performer's performance is continuously out of tune.

TECHNIQUE/ARTICULATION The performer demonstrates: *(Check ALL that APPLY, worth 2 points each)*

- ____ Appropriate and accurate tonguing.
- ____ Appropriate slurs as marked.
- ____ Appropriate accents as marked.
- ____ Appropriate ornamentation as marked.
- ____ Appropriate length of notes as marked (i.e., legato, staccato)

MELODIC ACCURACY The performer performs: *(Check ONE only)*

- 10 ____ all pitches/notes accurately.
- 8 ____ most pitches/notes accurately.
- 6 ____ many pitches accurately.

- 4 ____ numerous inaccurate pitches/notes.
 2 ____ inaccurate pitches/notes throughout the music, (i.e., missing key signatures, accidentals, etc.).

RHYTHMIC ACCURACY The performer performs: *(Check ONE only)*

- 10 ____ accurate rhythms throughout.
 8 ____ nearly accurate rhythms, but lacks precise interpretation of some rhythm patterns.
 6 ____ many rhythmic patterns accurately, but some lack precision (approximation of rhythm pattern used).
 4 ____ many rhythmic patterns incorrectly or inconsistently.
 2 ____ most rhythmic patterns incorrectly.

TEMPO The performer's tempo: *(Check ONE only)*

- 10 ____ is accurate and consistent with printed tempo markings.
 8 ____ approaches the printed tempo markings, yet the performed tempo does not detract significantly from the performance.
 6 ____ is different from the printed tempo marking(s), resulting in inappropriate tempo(s) for the selection, yet remains consistent.
 4 ____ is inconsistent (i.e., rushing, dragging, inaccurate tempo changes).
 2 ____ is not accurate or consistent.

INTERPRETATION The performer demonstrates: *(Check ONE only)*

- 10 ____ the highest level of musicality including well-shaped phrases and dynamics.
 8 ____ a high level of musicality, but has some phrases or dynamics that are not consistent with the overall level of expression.
 6 ____ a moderate level of musicality and musical understanding.
 4 ____ only a limited amount of musicality and music understanding.
 2 ____ a lack of musical understanding.

TOTAL SCORE: ____ / 70 POSSIBLE. *Please write this number in the space provided at the top.*

REFERENCES

- Aitchison, R. A. (1995). The effects of self-evaluation techniques on the musical performance, self-evaluation accuracy, motivation, and self esteem of middle school instrumental music students (Doctoral dissertation, University of Iowa). Dissertation Abstracts International, 56-IOA, 3875.
- Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education*, 41(1), 19-27.
- Bergee, M. J. (1997). Relationships among faculty, peer, and self-evaluation of applied performances. *Journal of Research in Music Education*, 45(4), 601-612.
- Brittin, R. V. (2002). Instrumentalists' assessment of solo performances with compact disc, piano, or no accompaniment. *Journal of Research in Music Education*, 50(1), 63-74.
- Byo, J. L. & Brooks, R. (1994). A comparison of junior high musicians' and music educators' performance evaluations of instrumental music. *Contributions to Music Education*, 21, 26-38.
- Fiske, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education*, 23(3), 186-196.
- Fiske, H. E. (1977). Relationship of selected factors in trumpet performance adjudication reliability. *Journal of Research in Music Education*, 25(4), 256-263.
- Hewitt, M. P. (2001). The effects of modeling, self-evaluation, and self-listening on junior high instrumentalists' music performance. *Journal of Research in Music Education*, 49(4), 307-322.
- Hewitt, M. P. (2002). Self-evaluation tendencies of junior high instrumentalists. *Journal of Research in Music Education*, 50(3), 215-226.
- Hewitt, M. P. (2005). Self-evaluation accuracy among high school and middle school instrumentalists. *Journal of Research in Music Education*, 53(2), 148-161.
- Hewitt, M. P. (2007). Influence of primary performance instrument and education level on music performance evaluation. *Journal of Research in Music Education*, 55(1), 18-30.
- Hewitt, M. P. & Smith, B. P. (2004). The influence of teaching-career level and primary performance instrument on the assessment of music performance. *Journal of Research in Music Education*, 52(4), 314-327.
- Lillis, G. (2000). Secondary instructional strategies: Education 452 syllabus. Cornerstone University, Grand Rapids, MI.

- McCoy, C. W. (1991). Grading students in performing groups: A comparison of principals' recommendations with directors' practices. *Journal of Research in Music Education*, 39(3), 181-190.
- Music Educators National Conference. (1994). The School Music Program: A New Vision. Retrieved December 3, 2007, from <http://www.menc.org/publication/books/prek12st.html>
- Morrison, S. J., Montemayor M., & Wiltshire, E. S. (2004). The effect of a recorded model on band students' performance self-evaluations, achievement, and attitude. *Journal of Research in Music Education*, 52(2), 116-129.
- 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(2), 259-272.
- Stegman, S. F. (2009). Michigan state adjudicated choral festivals: Revising the adjudication process. *Music Educators Journal*, 95(4), 62-66.
- U.S. Department of Education. (2002). No Child Left Behind. Retrieved December 3, 2007, from www.ed.gov/nclb
- Wells, R. (1997). Designing Curricula Based on the Standards. *Music Educators Journal*, 84(1), 34-39.

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 03063 3865