EXPLORING THE RELATIONSHIP BETWEEN APPLIED CLASSROOM ENSEMBLE FUNDAMENTALS AND ENSEMBLE PERFORMANCE ACHIEVEMENT

by

CHRISTOPHER DOBBINS

(Under the Direction of Brian Wesolowski)

ABSTRACT

The purpose of this study was to explore the relationship between applied classroom ensemble fundamentals and ensemble performance achievement of high school bands. This study was guided by the following research questions: (a) is there an overall relationship between exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement?; (b) what is the relationship between exposure to (e.g., time and type) varying types of ensemble-based fundamentals and ensemble performance achievement?; and (c) how well can performance achievement be predicted based upon exposure to (e.g., time and type) ensemble-based fundamentals. Thirty-four high school band directors were surveyed using the Ensemble Fundamentals Survey and the same high school band directors submitted single recordings taken from formal performance assessments. The recordings were blindly rated by subject matter experts (N=35). Observed rater data was converted to linear performance achievement measures using the Multi-Facet Rasch Partial Credit Measurement Model. These results were correlated with individual response items from the Ensemble Fundamental Survey in order to observe relationships between applied ensemble-based fundamentals and performance achievement. Results indicated that there was an overarching relationship between applied

ensemble-based fundamentals and performance achievement, that significant relationships were negative in direction, and that at least one ensemble-based fundamental (balance and blend exercises) had a negative predictive relationship with ensemble performance achievement. Discussion includes implications of the findings and suggestions for future research in the area of the relationship between ensemble-based fundamentals and ensemble performance achievement.

INDEX WORDS: band rating rubric, ensemble-based fundamentals, optimal scale regression, performance achievement, Rasch, Spearman's rho

EXPLORING THE RELATIONSHIP BETWEEN APPLIED CLASSROOM ENSEMBLE FUNDAMENTALS AND ENSEMBLE PERFORMANCE ACHIEVEMENT

by

CHRISTOPHER DOBBINS

BM & BME, Hastings College, 2007

MM, The University of Utah, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF EDUCATION

ATHENS, GEORGIA

2017

EXPLORING THE RELATIONSHIP BETWEEN APPLIED CLASSROOM ENSEMBLE FUNDAMENTALS AND ENSEMBLE PERFORMANCE ACHIEVEMENT

by

CHRISTOPHER DOBBINS

Major Professor:	Brian Wesolowski
Committee:	Clinton Taylor
	David Haas

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

DEDICATION

This document is dedicated to my wife, Heather, and daughter, Ginevieve. Without their love, support, and sacrifice, I would never have been able to begin this endeavor, let alone finish. This document is further dedicated to my parents, Jeff and Janet, who instilled in me a love of learning and from a young age supported and encouraged my exploration of various aspects of music and art.

ACKNOWLEDGEMENTS

I wish to profusely thank my doctoral committee. Thanks to the help and support of my adviser, Dr. Brian Wesolowski, I was confident enough to pursue as involved and valuable of a project as this document. His constant help and encouragement throughout the process was invaluable. Thanks to Dr. Skip Taylor for providing incredibly helpful commentary and advice on numerous aspects of this document and for good humor throughout the process. Thanks to Dr. David Haas for being a steady positive presence on my committee and stepping in at a difficult time to fill a vacancy, thereby allowing me to continue and progress with my studies. I also wish to thank Dr. Johanna Royo for her help in navigating bureaucratic logistics with the Graduate School. Very special thanks to Ms. Susan LeCroy, without whose help from even before I entered the program I would have been completely lost. Lastly, thanks to my colleagues at Washington and Lee University, especially Dean Suzanne Keen and Dr. Gregory Parker, for placing trust in me to lead the instrumental music program and finish this project simultaneously.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Ensemble-based fundamentals are exercises performed by an entire band, typically at the beginning of a rehearsal, which are designed to help the ensemble improve in specific aspects of band performance (e.g, tone production, rhythm and pulse accuracy, pitch and intonation accuracy, and expressive qualities) (Garofalo, 1973; Garofalo & Whaley, 1979; Love, 1997). Proponents of the consistent use of ensemble-based fundamental practice, such as Garofalo (1973), Younger (1998), Brophy (1979), and Manfredo (2006) argue that consistent fundamental practice is the best way to help a band play consistently in time, in tune, and with a good sound. Those who do not promote the consistent use of ensemble-based fundamental practice, such as Worth (2006) and Umberson (1970), argue that using ensemble time on fundamental practice takes away from music-making and can lead to harried, underprepared performances that yield less than desirable results and give students an overall experience that is less than positive.

This introduction will discuss reasons for both including and omitting the practice of ensemble-based fundamentals in the normal rehearsal process. It will then speak to both arguments' ties to assessments with both an ensemble contest/assessment (e.g., district music contest, large group performance assessment, etc.) and holistic (e.g., day-to-day student progress and overall educational growth assessments) viewpoint. Finally, it will focus on student learning outcomes in terms of performance achievement, and why exploring the relationship of ensemble-based fundamentals and performance achievement is important.

Ensemble-based Fundamentals in Rehearsal Settings

Proponents for the use of ensemble-based fundamental practice defend their argument with pedagogical ideas drawn from the research literature. Fundamental practice is an integral part of a well-paced rehearsal (Brendell, 1996), and pacing can play an important role in consistent ensemble development (Blocher, Greenwood, & Shellahamer, 1997). By practicing fundamentals during the band rehearsal, students are able to learn the component parts (e.g., notes, rhythms, dynamics, articulations) of music for performance more quickly (Manfredo, 2006). Efficiency in learning helps students to play more difficult music more accurately (e.g., fewer incorrect notes, rhythms, dynamics, articulations) and frees them to perform in a more expressive way while taking less overall rehearsal time to do so (Garofalo, 1973; Garofalo & Whaley, 1979; Manfredo, 2006). The types of ensemble-based fundamentals used to achieve the efficiency and skill outcomes cited by Garofalo (1973), Garofalo and Whaley (1979), and Manfredo (2006) is varied. Love (1997) found that high school band directors employ some combination of long tones, lip slurs, scales, method/technique books, tuning, chorales, original exercises, sight-reading, listening exercises, rhythm drills, articulation exercises, breathing exercises, dynamic exercises, ear-training, exercises to help achieve balance/blend, and singing exercises in time allotted for ensemble-based fundamental practice. While it is impossible to cover each of these aspects of ensemble-based fundamental practice in every rehearsal (Bauer, 1993; Manfredo, 2006; Umberson, 1970; Worthy, 2006; Yarbrough & Price, 1989), Love (1997) found that most high school band directors who used ensemble-based fundamentals on a regular basis would focus on a few individual techniques per day. For example, a high school band director might focus on long tones, scales and lip slurs one day, use balance/blend exercises and breathing exercises the next, and continue to alternate and adjust the ensemble-based

fundamental exercises used based on the aspect of performance the band needed the most help with at that particular time (Garofalo, 1973; Love, 1997; Manfredo, 2006). Most high school band directors who use ensemble-based fundamentals would then move on to the music being studied for performance (Bauer, 1993; Love, 1997; Manfredo, 2006; Worthy, 2006; Younger, 1998) While the amount of time spent on fundamental practice varies significantly based on director preference (Love, 1997; Manfredo, 2006), evidence exists that less experienced teachers use less than half as much time as experienced teachers on performance activities encompassing ensemble-based fundamentals (Goolsby, 1996).

Efficient practice of ensemble-based fundamentals can be part of a well-timed and well-paced rehearsal (Goolsby, 1996; Manfredo, 2006). Manfredo (2006) argues that pacing is the single biggest contributor to holding an efficient and productive rehearsal. The numerous component parts of a rehearsal (setup, announcements, ensemble-based fundamental practice, literature rehearsal, tear-down, etc) as cited by Manfredo (2006) can put a time strain on bands preparing for contests or assessments and can lead to the omission of ensemble-based fundamentals as a teaching tool (Opsal, 2013). It is largely due to the time strain placed on high school bands and their directors that a number of high school band directors use no ensemble-based fundamental practice in a normal rehearsal (Opsal, 2013). Some high school band directors omit ensemble-based fundamentals, but do use different parts of the literature being studied to teach fundamental aspects of music that could otherwise be taught through ensemble-based fundamental exercises (Banister, 1992). The heavy emphasis on contest/assessment ratings and the results of those activities increasing use in teacher evaluations cause some high school band directors to forgo ensemble-based fundamental practice altogether (Austin, 1990) or try to move it into the solo/small ensemble portion of the band program, if one exists (Meyers, 2011).

Contests/Assessments versus Holistic Use

Different types of contests/assessments have been developed as the emphasis on participation in contests/assessments has increased for high school bands (Saidon & Shah, 2013). Some contests and assessments are non-competitive and strictly for comments or unposted ratings (Saidon & Shah, 2013). These events can be structured in a way that is helpful to directors in adjusting curriculum and pedagogy based on performance outcomes that can help ensembles improve over time (Saidon & Shah, 2013). Other events are competitive and end with a ranking of highest to lowest performance achievement or with posted ratings as assigned by the adjudicators (Saidon & Shah, 2013). These events can be helpful to curriculum development, but can also hinder a directors ability to motivate their students (Sheldon, 1994). Ranked competitions can sometimes result in the music education version of "teaching to the test" (Austin, 1990) as directors strive for more ensemble-based competitive success rather than individual student musical growth.

Emphasis on "official" contests/assessments (events organized and run by state, district, or other agencies with official capacity) has, in some cases, caused preparation for these events to replace more holistic assessment in high school band classrooms (McCoy, 1991; McPherson & Thompson, 1998). Assessment as a whole should be geared more towards the improvement of the student musician and the ensemble experience, rather than perceived success in one-off events (Austin, 1990; Banister, 1992; McCoy, 1991). The practice of ensemble-based fundamentals can aid in individual student improvement (Garofalo & Whaley, 1979; Maynard, 2006), but does take up portions of rehearsal that could be spent on literature for performance (Manfredo, 2006). The value of consistently practicing ensemble-based fundamentals is

undefined, however, as the relationship of ensemble-based fundamentals to performance achievement has not been thoroughly explored.

Research Questions

The purpose of this study is to examine the relationship between the practice of ensemble-based fundamentals and ensemble performance achievement. The study was guided by the following research questions:

(a) Is there an overall relationship between exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement?

(b) What is the relationship between exposure to (e.g., time and type) varying types of ensemble-based fundamentals and ensemble performance achievement?

(c) How well can performance achievement be predicted based upon exposure to (e.g., time and type) ensemble-based fundamentals?

CHAPTER 2

REVIEW OF LITERATURE

Numerous factors contribute to the effectiveness of the practice of ensemble-based fundamentals in the high school band classroom (Garofalo, 1973; Garofalo & Whaley, 1979; Manfredo, 2006). A review of the literature reveals that there are numerous types of ensemble-based fundamentals and different ways to use them, many non-musical factors that can potentially influence the effectiveness of ensemble-based fundamental practice, and different ways to formally and informally assess the effectiveness of the ensemble-based fundamentals that are practiced. This literature review addresses each of those issues, respectively, and closes with a discussion of the psychometric considerations that will be used in the analysis of data.

Use of Ensemble-Based Fundamentals in Rehearsal

A sound curriculum with ensemble-based fundamental practice at its core is the most efficient way to improve overall band performance (Garofalo, 1973). Garofalo (1973) pushes for sound curriculum that consists of not just the score being studied, but fundamental exercises outside of music for performance. Garofalo (1973) does not suggest specific exercises, but does point to ensemble-based fundamental practice as being important. Garofalo and Whaley (1979) suggests developing a unit-based approach to curriculum design that includes ensemble-based fundamental practice at its core. The ensemble-based fundamentals chosen to be consistently studied throughout the unit should be derived from aspects of the music being studied which the students cannot consistently execute. Developing curriculum this way, as opposed to using the

score as the curriculum, leads to higher levels of musicianship and higher quality ensemble performance achievement (Garofalo & Whaley, 1979).

Love (1997) describes the most commonly used ensemble-based fundamentals as long tones, lip slurs, scales, exercises from method/technique books, tuning, chorales, original fundamental exercises, sight-reading, listening exercises, rhythm drills, articulation exercises, breathing exercises, dynamic exercises, ear-training, exercises to help achieve balance/blend, and singing exercises. It is noted that it is likely impossible to incorporate each of these varied ensemble-based fundamentals into every rehearsal (Love, 1997; Manfredo, 2006). Therefore, the type of ensemble-based fundamentals most commonly performed and the time spent on those fundamentals play a key role in their effect on ensemble performance achievement (Love, 1997). Love's work further validates aspects of what Bergee (1992) deemed important in his work evaluating the effectiveness of student teachers. In that study, Bergee (1992) noted that student teachers who broke rehearsals down into sections, including a section for warm-up/fundamental activity, had more success than those student teachers who focused only on music for performance.

The study of fundamentals for the sake of studying them does not necessarily lead to higher quality ensemble performance achievement (Duke & Simmons, 2006). The types of fundamentals performed, and most importantly a focus on the sound produced (Duke & Simmons, 2006) has a positive relationship with ensemble performance achievement. While simply going through the motions seems to be better than nothing, intentional focus on the quality of sound drives improvement (Bauer, 1993; Duke & Simmons, 2006; Yarbrough & Price, 1989; Younger, 1998). Cavitt (2004) discovered that band directors spend much rehearsal time on the correction of intonation errors. In her study, she noted that daily time spent on ensemble-

based fundamentals had a relationship with better overall intonation, even if the fundamentals practiced were not specifically oriented to improve intonation. Cavitt (2004) describes the repetition of fundamentals as being vital to the ear training of band students, and notes that students who could tune using ears rather than with an electronic tuner were able to play with higher levels of correct intonation. She attributes this, in part, to the repetition of ensemble-based fundamentals and the insistence of band directors to repeat ensemble-based fundamentals with an ensemble-wide focus on improved intonation, even if the specific fundamental being practiced might have a purpose that lies outside of intonation study (Cavitt, 2004).

Repetition of musical activity and musical phrases is an important part of any practice session, and can lead to deeper learning with less use of time (Maynard, 2006). This finding, when applied to the practice of ensemble-based fundamentals, suggests that picking fewer fundamental activities that are deemed more important for pedagogical purposes and performing those activities more often will lead to higher quality performance achievement (Garofalo & Whaley, 1979; Maynard, 2006). In conjunction with repetition, incorporating musicianship into ensemble-based fundamental activity may lead to higher quality ensemble performance achievement (Paschall, 2006). Paschall (2006), in a review of beginning band method books, noted that the most popular ensemble-based fundamentals are too often performed with no embedded musicianship. Embedded musicianship is the performance of a musical task including expressive qualities and stylistic interpretation, rather than just focusing on notes and rhythms (Paschall, 2006). Fundamentals for fundamentals sake, then, may not be conducive to overall musicianship and could lead to lackluster performance. Bergee (1992) notes that effective beginning teachers embed musicianship into their practice of ensemble-based fundamentals, and

that their rehearsals are more effective than their peers who do not embed musicianship into ensemble-based fundamentals or do not practice them at all.

Non-Musical Factors Influencing Effectiveness of Ensemble-Based Fundamentals

A review of the literature reveals several non-musical factors that have the potential to influence the effectiveness of ensemble-based fundamental practice. These factors include rehearsal structure, rehearsal pacing, teacher/director activity and influence, and student motivation.

The relationship between rehearsal structure and ensemble performance achievement was discussed by Bauer (1993). He found that time spent on balance between sections, overall band intonation, and practice counting difficult rhythms aloud (all ensemble-based fundamentals as described above and codified by Love (1997)) all have a significant relationship to higher quality performance achievement. The way in which a rehearsal begins can affect the quality and efficiency of the rehearsal as a whole. Brendell (1996) notes that activities which do not engage students in musical activities can lead to higher instances of off-task behavior. Where off-task behavior is higher, time is used less efficiently (Brendell, 1996; Manfredo, 2006). A focus on efficient use of time to increase on-task behavior is important in increasing ensemble performance achievement (Brendell, 1996). Manfredo (2006) lays out ways in which a rehearsal can be organized in order to be most efficient and productive. He does not speak directly to the use of ensemble-based fundamentals in each rehearsal, grouping them into a "warm-up" period, but does argue that efficient use of time leads to higher quality ensemble performance achievement (Manfredo, 2006). In addition to carefully planning how time is spent when incorporating ensemble-based fundamentals into the overall rehearsal plan, directors who focus on the performance of fundamentals rather than on teacher talk tend to experience higher quality

ensemble performance achievement (Witt, 1986). Therefore, focusing on "performing" fundamentals and spending less time with wordy verbal instruction helps to increase time on task and makes the practice of ensemble-based fundamentals more valuable (Price, 1983; Witt, 1986; Worthy, 2006). This does not mean more overall time spent on ensemble-based fundamentals, but rather that the time set aside for ensemble-based fundamental practice should be filled with performance rather than verbal instruction (Witt, 1986). Goolsby (1996) confirms Witt's assertion. He found that teachers who teach ensembles known to present high quality performances were teachers who talked the least during rehearsals. In addition, these teachers used nonverbal modeling, spent more time on performance than on practice, and divided rehearsals into easily recognizable sections, one of which was used for warm-up and ensemble-based fundamental activities (Goolsby, 1999).

No matter what activities take place over the course of a normal rehearsal, how those activities are paced can play an important role in ensemble performance achievement (Duke, Prickett, & Jellison, 1998). Where ensemble-based fundamentals are concerned, the place in which they are performed in the rehearsal (beginning, middle, or end), how much time is spent on them, and the ratio between student performance time and teacher talk time all contribute to the overall effectiveness of the ensemble-based fundamental activity on performance achievement (Bauer, 1993; Blocher et al., 1997; Brophy, 1979; Gillis, 2008; Manfredo, 2006; Price, 1983; Yarbrough & Price, 1981, 1989). Younger (1998) argues that pacing could play an important role in how effective the practice of ensemble-based fundamentals ends up being. While some believe that there is much room for experimentation in setup, planning and pacing (Umberson, 1970), a consistent rehearsal order can lead to less need for teacher talk, thereby increasing student performance time and eventual overall effectiveness of the ensemble-based

fundamental activity (Brendell, 1996; Duke et al., 1998; Goolsby, 1996; Witt, 1986). Brophy (1979) mentioned higher quality ensemble performance achievement as a possible outcome of consistently well-paced rehearsals.

The behavior or activity of the teacher/director can play a part in the success of the practice of ensemble-based fundamentals (Blocher et al., 1997; Brophy, 1979; Price, 1983; Younger, 1998). Directors who engage in conceptual teaching of fundamental concepts in order to aid transfer tend to have higher levels of concept retention that can lead to performances being perceived as higher quality (Blocher et al., 1997). Similarly, a director who is high magnitude (consistently energetic, enthusiastic, and demanding) can lead to less off-task time and higher quality performance achievement when compared to low magnitude directors (Yarbrough, 1975). Teacher intensity can play a positive role in student engagement with difficult concepts (Cassidy, 1990; Silveira, 2014). Teachers with certain personality types may also engage students in a way that leads to more on-task time with higher levels of motivation, which leads to higher quality ensemble performance achievement (Teachout, 2001). It is also the conductor's responsibility to be as prepared as possible in order to be able to teach high level musical concepts (such as comprehensive musicianship) and not just notes and rhythms (Gillis, 2008). How a future teacher is trained to develop lesson (or rehearsal) plans can play a part in how effective and efficient a rehearsal they are able to run (Lane, 2010). A prepared teacher is less likely to have large stoppages of time in which they are not giving instructions or engaging the ensemble in performance (Goolsby, 1999). The amount of information presented in any stoppage of performance and how that information is presented can play a large role in effecting performance achievement (Worthy, 2006). Information that is presented to the group in a process that follows presentation of the academic task, instructions on execution, performance, and instructor

feedback show the biggest gains when compared to other models (Price, 1983). Having a consistently high level of expectation and demands can also lead to higher quality ensemble performance achievement (Yarbrough & Price, 1989). Engaging students at a high level with only short stoppages for instructions that address tone, balance, intonation, and other musical traits rather than just notes and rhythms are one sign of an effective and experienced teacher (Goolsby, 1999).

Student motivation also plays a role in determining how successful the practice of ensemble-based fundamentals is for a high school band (Opsal, 2013). If students are motivated, whether by the music being studied, the director, or an outside influence, they will focus harder and perform at a higher level (McPherson & Hendricks, 2010). Motivation can come in many different forms. Love of the music being studied leads to high levels of student motivation (McPherson & Hendricks, 2010). Interest in specific pieces or time periods makes students more interested in performing those pieces at a higher level. This motivation serves to help them focus on little things, such as the transfer of ensemble-based fundamental activity to full pieces, which in turn leads to higher quality performances (McPherson & Hendricks, 2010). The director/teacher themselves can provide motivation to students (McPherson & Hendricks, 2010). It has been noted that master teachers find ways to extrinsically motivate students in a way that breeds intrinsic motivation over time (Worthy, 2006). Sometimes personality plays a factor in how motivationally effective a director is (Teachout, 2001), but more often director driven motivation stems from the magnitude of their behavior (Yarbrough, 1975) and how well they are able to maintain the attention of the ensemble through performance activity (Yarbrough & Price, 1981). Purely extrinsic factors can also motivate students to take rehearsal time, and time spent practicing ensemble-based fundamentals, more seriously (McPherson & Hendricks, 2010).

Competition is perhaps the most ubiquitous way in which music students are motivated to excel (Opsal, 2013). While not inherently negative, the use of competition as motivation can lead to the music education version of "teaching to the test", and result in music contests or assessments essentially becoming high stakes testing (Saidon & Shah, 2013). There is danger here in losing the art behind studying music (Opsal, 2013). In some instances, teacher effectiveness is also being tied to contest rating, furthering the emphasis on earning top ratings at contests and potentially hurting individual student artistic achievement and motivation for higher performance achievement (Opsal, 2013).

Assessment Types and Tools

There are numerous ways in which music educators assess student achievement. The most important factor in assessment is developing a reliable and valid model for assessment (Asmus, 1999). Properly gathered, assessment information is valuable to teachers and students alike and can help to drive the improvement of the individual student and the ensemble in which s/he performs. Unfortunately, much of the talk about music assessment is related to various music contests as opposed to a more wholistic, student-centered models (Bergee, 2007). Band contests do not take into account individual student improvement or achievement, and therefore do not show then whole picture of performance achievement in relation to ensemble performance (Saidon & Shah, 2013). However, band contests/assessments have become an increasingly important way of determining the overall success of a band program and influence how teachers teach (Banister, 1992; Saidon & Shah, 2013; Sheldon, 1994).

Band contests, assessments, and festivals have evolved to take many different forms. There are competitive and non-competitive events, those sanctioned by state or regional organizations, and those put on by individual ensembles or schools (Saidon & Shah, 2013). No

matter the type of event, the end goal is the same: To find out how one's band stacks up against others through comments, and/or ratings, and/or rankings. This emphasis on comparison plays a role in how students and those around them (parents, administrators, etc) view ensemble performance achievement (Sheldon, 1994). In some cases, this even extends to the point of a band that earns a second division rating at a contest, despite that specific performance being their best of the season, feeling less positive about that performance (Austin, 1990).

There are numerous contributing factors to the perceived quality of performance achievement at contests. Often, how the director or students view marching or concert band contests can have an impact on the score they earn and how that score effects their overall band program (Banister, 1992). The more a group can view contests as part of the process towards overall improvement, the better. Groups that view contest ratings at the end of a season as the final say on the quality of their ensemble often experience extreme highs and lows and end up with less improvement over the course of time (Banister, 1992). Some directors are able to use contest comments and ratings to help drive improvement (Austin, 1990; Saidon & Shah, 2013). Sometimes this takes the form of convincing students that extra-musical factors, such as field or performance space conditions and temperature had an undue positive or negative effect on their rating (Saidon & Shah, 2013). At other times, comments can serve as a necessary extrinsic motivator for students to work harder to achieve their ensemble goal of a high rating or ranking (Austin, 1990).

The way in which a conductor leads her/his ensemble during the run up to a performance can also have an effect on the performance quality of the ensemble (Morrison, Price, Geiger, & Cornacchio, 2009; Price & Chang, 2001, 2005a). While the expressivity of the conductor during the actual contest performance does not have a significant effect on performance quality (Price &

Chang, 2001), how a conductor prepares her/his ensemble to perform in an expressive way through fundamental practice can have a significant effect on ensemble performance achievement (Price & Chang, 2005). The attitude a director or her/his students have about contests and their importance or benefit can also have an effect on ratings. Meyers (2011) found that director attitudes towards contests could have a direct impact on the quality of their student's performances in those events. Those with more positive attitudes tended to have higher scores than those with negative attitudes (Meyers, 2011). Most directors believed there was a benefit to the students when they participated in contests (Meyers, 2011). The directors who could articulate what those benefits were tended to have higher ratings than those who could not (Meyers, 2011).

There are other factors that lead to diversity of scores earned at contests. Warm-up rooms, temperature, the ensemble that immediately preceded your ensemble, and rater attitudes can all have an effect on ratings at contests (Bergee & McWhirter, 2005). Further, ensembles with more experience practicing in near identical conditions to the contest and ensembles with a high percentage of students engaging in private applied study tend to have higher ratings (Hamann & Banister, 1991). The type of rating scale used in a contest or assessment can have a significant impact on how ratings are assigned, how consistent or inconsistent ratings are throughout a contest, and can even effect the raters themselves (Hash, 2012; Kinney, 2009; Wesolowski, Wind, & Engelhard, 2015, 2016a, 2016b). Typically, there are two types of rating scales used in music performance assessment: Likert-type rating scales, and criteria specific rating scales (Saunders & Holahan, 1997). Both types of scales can provide valuable information, but are imperfect in their assigning of ratings as they de-emphasize or exclude important facets of performance that should be accounted for in a holistic rating. As early as

1965, Gutsch (1965) was advocating for the development of tools that would measure students musical achievement in a global sense, rather than their factual knowledge of music or simply whether they were performing correct or incorrect notes and rhythms. Often, contest rating scales ask the adjudicator to rate the ensembles overall performance on some scale of 1-5, with 1 being a superior rating, and 5 being poor (Wesolowski, 2012). Some provide a rubric that may or may not have been developed with positive pedagogical intent, but many simply provide some general guidelines as to what should be listened for in each caption before assigning the overall rating. For reference, high school band adjudication forms from Southern California (Appendix A), Georgia (Appendix B), Iowa (Appendix C), Kansas (Appendix D), Nebraska (Appendix E), North Carolina (Appendix D), Texas (Appendix E), and Virginia (Appendix F) are included.

Generally, the use of a rubric as part of the assessment data helps to increase rating consistency (Asmus, 1999). Further, a well-developed rubric can "serve as documentation for student achievement that provides music teachers with a written form of accountability" (Wesolowski, 2012). Rubrics alone are not enough, however, as rater training and familiarity with the rating tool can also confound rating scores (Kinney, 2009). Criteria-specific rating scales can help make a 1-5 scale more reliable and consistent (Saunders & Holahan, 1997). While the rating given at the end of the assessment is often still couched in terms of 1-5, a breakdown of expected and acceptable performance characteristics can help provide more consistent ratings (Saunders & Holahan, 1997). However, these rating scales do not account for any type of overall quality judgement or statement about achievement level; they merely serve as an accounting checklist of performance criteria that are either present or absent (Wesolowski, 2012). Latimer et al (Latimer Jr., Bergee, & Cohen, 2010) undertook a large study of the effectiveness and perceived pedagogical utility of an assessment rubric used in Kansas for music

performance assessments(Latimer Jr. et al., 2010). This rubric was weighted, and its proper use allowed adjudicators better justification for ratings assigned. While its rating reliability was similar to other tools researched, the addition of the multidimensional rubric yielded comments pointing to adjudicators and directors perceiving the rubric to possess improved pedagogical utility (Latimer Jr. et al., 2010). While the tool did not prove more reliable than past tools, the increased confidence by those using the tool and deciphering its results is a positive step. An example of the rubric studied by Latimer and his colleagues is included as part of Appendix D (Latimer Jr. et al., 2010).

Even when using rubrics that are developed to be multidimensional and to help adjudicators and directors are used in contests/assessments, the qualities that are focused on are sometimes difficult to perceive. For example, Bergee (1995) found that important rubric items such as intonation, balance, blend, and tone quality were often inextricably linked. This means that rubric items pointed at those parts of performance are likely to be unreliable and could confound overall ratings or scores. In a study examining the reliability of two choral rating forms, Norris and Borst (2007) found results that further validate above studies. They used a traditional rating form with no instructions or rubric, and compared it to a second rating form that included instructions to adjudicators by means of a rubric. They found that the second rubric-oriented form was more reliable and a better assessment of ensemble performance achievement, but that several rubric items were linked and could therefore confound results (Norris & Borst, 2007).

Using the facet factorial approach, an approach which uses weighted facets of performance as factors in the final rating or analysis, DCamp developed the *Band Performance Rating Scale (BPRS)*, which requires the adjudicator to respond to thirty unique items related to

the band performance being judged (DCamp, 1980). The *BPRS* showed high reliability and validity in a test-retest setting when scores from all seventeen judges were included (DCamp, 1980). It is noted, however, that the reliability and validity may have been higher because of the relatively large sample size for adjudication (DCamp, 1980). Typically, three judges are employed to rate bands at festivals or contests (Hash, 2013; Kinney, 2009). The smaller number could create higher levels of variability, making the tool less reliable and valid (DCamp, 1980). DCamp (1980) suggests that training of raters using the tool could alleviate this variability and further argues that the specific feedback provided through use of the tool is of greater value to individual directors. DCamp's *BPRS* in its original form is included in Appendix I.

Official band contests/assessments (those planned and run by state, district, or other agencies) use adjudicators to help to rate bands. Subject matter experts are often used and are identified by characteristics which may include (but are not limited to) years of service/experience, success as a music educator, and their ability to identify, diagnose, and give solutions to common problems experienced during performance (Kruth, 1970). Over the course of time, band ratings at contests/assessments have steadily risen (Boeckman, 2002). In most 1-5 rating systems, there seems to be a preponderance of 1-3 ratings, with ratings 4 and 5 rarely being used (Boeckman, 2002). One reason for this inflation is that there is a reluctance on the part of raters to assign a rating lower than 2 (Boeckman, 2002). Since the 1970's, ratings have steadily risen, resulting in nearly half of all ratings being either a 1 or a 2 (Boeckman, 2002). While some of this is likely due to bands overall having improved, much if it is due to ratings inflation and rater unreliability (Boeckman, 2002). Some researchers have shown that, under the right circumstances, rater reliability and inter-rater reliability of music performance assessment can exhibit good reliability (Bergee, 2003). This validity and reliability is largely due to

adjudicators using a well-developed tool and receiving training or having enough experience in adjudication (i.e. conducting end-of-semester juries for a number of years) to help reduce potential confounds (Bergee, 2003; Iusca, 2014; Kinney, 2009). The more the adjudicator is invested in overall student success, the more reliable the rating as well, as the rater will be more willing to assign low but honest marks in an effort to be transparent with the student. (Bergee, 2003). Proper training of adjudicators on the contest format and contest tool is the best way to increase rater reliability and validity and eliminate potential confounds involving raters (Bergee, 2007). This is important, as Bergee (2007) found that raters are often the greatest source of variability in a given contest or assessment, and as such are often responsible for unreliable results.

The tool itself is often a problem in reliability of ratings. In a study of the reliability of ratings at the 2005 Virginia Band and Orchestra Directors Association State Marching Band Festival, King and Burnsed (2009) found a high level of inter-rater reliability throughout the event. Upon further examination, they posit that this had more to do with the tool not differentiating enough between performance captions and thereby leading adjudicators to higher scores through lack of specificity regarding expected performance outcomes for each rating category (King & Burnsed, 2009). In this case, inter-rater reliability was high, but because of an inadequate tool, the results were still inflated (King & Burnsed, 2009). Various ratings tools take many different forms, with some tools being less reliable than others (Burnsed, Hinkle, & King, 1985). We have already seen the standard 1-5 (or Superior-Poor) is inadequate from a reliability standpoint. As early as 1965, music educators were searching for ways in which a tool could be created that would be reliable and valid in evaluating parts of instrumental music performance (Gutsch, 1965). While developed specifically to evaluate sight-reading, Gutsch's (1965) scale

proved reliable and valid, but was also too cumbersome to use in a true contest setting. Increasing reliability can be as simple as adding descriptive extensions to an already utilized tool (Norris & Borst, 2007). This allows directors to continue to use the bottom-line score that they are used to, but provides more information and creates higher levels of interrater reliability and accurate score prediction (Norris & Borst, 2007). Similarly, creating a weighted rubric with pedagogical reasons for weighting each section can help to make tools more reliable (Latimer Jr. et al., 2010). In this instance, not only were the ratings found to be within range of other descriptive tools in terms of reliability, a post-contest survey of directors found that the group surveyed perceived the rubric as having improved pedagogical utility (Latimer Jr. et al., 2010).

A technique that has been tried with varying success is the use of a Likert-type polytomous rating scale to help in the assessment process. While these scales often show reliability, their validity can be called into question (Ciorba & Smith, 2009). Additionally, Likert-type scales ask for a level of agreement with a general statement, and therefore may not reflect the adjudicators true perception of performance (Wesolowski, 2012). The results taken from a Likert-type scale can also be hard to interpret, potentially making the results less useful for directors and students (Engelhard, 2002; Wesolowski, 2012). This is often due to the fact that ratings are often associated with rater characteristics rather than the performances being rated (Engelhard, 2002). Criteria-specific rating scales have been used with substantial reliability (Saunders & Holahan, 1997). In this model, the adjudicator selects the statement that best describes several different categories of the performance being judged. While yielding results that are highly reliable, the criteria-specific scale is not yet perfect (Saunders & Holahan, 1997). There is still a lot of human error possible, and some of the criteria can be somewhat vague, leading to a penchant for the adjudicator to select the most neutral statement (McPherson &

Thompson, 1998; Saunders & Holahan, 1997). McPherson and Thompson (1998) showed that characteristic checklists, when included in criteria-specific rating scales, can help make results more reliable and valid. Still, there are too many options of characteristics to include, and so construction of the checklist can have a broad impact on ratings outcomes.

A potential next step has been to take the best of each rating system, and include some factorial analysis in order to mitigate the human element (Bergee, 1995). Separating factors, however, can be a difficult process. For example, it has been found to be difficult to accurately separate factors of balance, blend, tone quality, and intonation (Bergee, 1995). This can lead to confounded results when adjudicators are asked to separate these factors for analysis (Bergee, 1995). DCamp (1980) used an approach that included weighting of factors as well as statistical correction in order to correct for rater and selection severity. While an improvement over previous ratings tools, the approach requires significant training in order to be properly used by adjudicators, and requires more involved math in order to determine a final rating that can be expressed in the common 1-5 scale (DCamp, 1980). DCamp (1980) did develop a comprehensive list of ensemble fundamentals as part of his work, and that list was used to aid in development of the Ensemble Fundamentals Survey used as part of this study.

Perhaps the most frustrating part of ratings scales used to rate music assessments is that no matter the tool, adjudicator, or that adjudicators training, there is some amount of subjectivity (Boeckman, 2002; King & Burnsed, 2009; Kinney, 2009). Perhaps the best way to ameliorate subjectivity is to ensure that adjudicators have clear expectations for each rating that they may assign (Wesolowski, 2012). Rubrics are the easiest and most efficient way to accomplish this task. While rubrics are not perfect, they do increase reliability and validity of ratings and show consistency over time and between contests (Wesolowski, 2012). Further, they are easily

modified for unique circumstances or to place emphasis on different aspects of performance (Wesolowski, 2012). Rubrics also give clear feedback to directors and can help lead to more efficient improvement by ensembles rated using them (Wesolowski, 2012).

Psychometric Considerations

For many years, the main measurement model in music education research centered around rater-mediated assessment data has been Classical Test Theory (CTT). CTT utilizes the analysis of raw data in order to draw conclusions based on common statistical processes. CTT has been used to measure teaching effectiveness (Bergee, 1992), musical environment (Brand, 1985), musical affect (Asmus, 1985; J. S. Edwards & Edwards, 1971; Sandstrom & Russo, 2013; Shaw & Tomcala, 1976), musical aptitude (Asmus, 1989), and performance achievement (Bergee, 1995; Nichols, 1991; Smith, 2009; Zdzinski & Barnes, 2002). CTT has also commonly been used when analyzing the reliability, validity, and accuracy of performance evaluations, assessments, and contest ratings, as well as reliability and validity of adjudicator scores and specific ratings tools (Bergee, 2003, 2007; Bergee & McWhirter, 2005; Bergee & Platt, 2003; Bergee & Westfall, 2005; Boeckman, 2002; Brakel, 2006; Burnsed et al., 1985; Ciorba & Smith, 2009; Conrad, 2003; Garman, Boyle, & DeCarbo, 1991; Hash, 2012; King & Burnsed, 2009; Kinney, 2009; Latimer Jr. et al., 2010; Morrison et al., 2009; Norris & Borst, 2007; Price & Chang, 2005b; Saunders & Holahan, 1997; Silvey, 2009). CTT uses raw scores gathered from a group of participants who are being tested on their success or failure on a number of individual items (Asmus, 1989). The CTT measurement model assumes that scores taken from a measure are comprised of two parts: (a) true score, and (b) measurement error (Wesolowski, 2012). The ability of an item to show differences between participants of various ability levels is measured statistically through Pearson product-moment correlation coefficients (Wesolowski, 2012). When

Likert or Likert-type rating scales are used, adjusted proportion-correct values (*p*-values) and correlation coefficients are utilized in order to determine the difficulty of individual items and the overall ability level of a participant (Wesolowski, 2012).

Using CTT to analyze performance assessment data has the disadvantage of relying upon the sample and test dependency of estimated person parameters and item parameters (Wesolowski et al., 2016b). This dependency limits the development of valid and reliable measures and prevents development of informed inferences rated to the ability of the participants and difficulty of individual items that go further than the context of the single assessment situation being tested (Wesolowski et al., 2016b).

The Many-Facet Rasch Model

When compared to CTT, the Rasch family of measurement models offers a more grounded theory (Wesolowski et al., 2016b). Rasch measurement theory (Rasch, 1960/1980) is often preferred in the measurement of latent traits and in the development of scales in the behavioral, social, and health sciences (Engelhard, 2013). The major benefit of the Rasch model is that invariant measurement is achieved when adequate fit to the model is observed (Wesolowski et al., 2016b). In terms of assessments, invariant measurement implies that participant outcomes are not influenced by the particular items used to measure them, and the outcomes are not influenced by raters whose job it is to use individual items to measure participant achievement (Wesolowski et al., 2016b). Rasch models use probabilistic distributions of participant responses as a logistic function of person and item parameters in order to define a latent trait (Wesolowski et al., 2016b). Whereas CTT uses raw scores directly in the analyses, raw scores are converted to a log-odds scale using a logistic transformation when using Rasch measurement theory in order to create a linear scale (Wesolowski et al., 2016b). The transformed

data can then be used as a dependent variable with multiple independent variables (or facets) that might be of interest (Wesolowski et al., 2016b). Examples of facets of interest might include rater severity and rater leniency, individual item difficulty, individual task difficulty, and performance achievement level (Wesolowski et al., 2016b). Each relevant item is placed into a hierarchy of difficulty, and each participant's individual item responses are mapped onto a single logit (log-odds units) scale. The result of the mapping of facets onto a single continuous variable scale makes possible the construction of a visual variable map to help illustrate the differences in difficulty among facets (Wesolowski et al., 2016b).

The Partial Credit version of the Rasch model (Masters, 1982) allows an additional interaction parameter for the analysis of individual rating scale categories within each separate item in the measure. This provides the basis for challenging the assumption that all items on a rating scale are spaced equally according to task difficulty (A. S. Edwards, 2017). This also allows for improvement in optimization and precision of the item scales being analyzed because it reviews the unique difficulties of each task encompassed in each item (A. S. Edwards, 2017). The statistical output of the Rasch analysis also allows for the review of category distribution within items and appropriate discrimination between performances (Linacre, 2002). The Rasch family of models provides not only a meaningful method for the development of valid and reliable measures, but provides a method for the consistent evaluation of rater quality within the context of performance-based assessments (A. S. Edwards, 2017).

CHAPTER 3

METHOD

Ensemble Fundamentals Survey

The Ensemble Fundamentals Survey (EFS) was developed based on the work of DCamp (1980) and Love (1997). A list of ensemble-based fundamental exercises most commonly used by high school band directors was defined by Love (1997). This list was used to define the ensemble-based fundamental options in the EFS which high school band directors were asked to choose from. The list of ensemble-based fundamental exercises is paired with options for frequency of use throughout rehearsals to further define time and type of exposure. Taking into account different scheduling, the participating high school band directors were given choices based on the proportion of rehearsals (e.g., every rehearsal, more than half of all rehearsals, less than half of all rehearsals, once per week or less, or never) in which they engage or do not engage in specific ensemble-based fundamental activities. The frequency periods which high school band directors were able to choose from were drawn from the work of Price (1983), Younger (1998), Garofalo &Whaley (1979), Gillis (2008), Lane (2010), Manfredo (2006), Bauer (1993), and Love (1997). The frequency periods were used to define frequency of exposure to specific ensemble-based fundamental activities. The list of ensemble-based fundamental activities was then paired with varying numbers of minutes (less than one minute, 1-3 minutes, 3-5 minutes, more than five minutes) during which the ensemble-based fundamentals are practiced during a given rehearsal. The time breakdowns are derived from the work of Manfredo (2006), Garofalo & Whaley (1979), and Love (1997). Determining the time each type of

ensemble-based fundamental is used during each frequency period is an important factor in the examination of the relationship of ensemble-based fundamentals to performance achievement outcomes.

The participating high school band directors were then asked to select the range of total time spent on ensemble-based fundamentals during a typical rehearsal. The times which participants were able to choose from were drawn from the work of Love (1997), Manfredo (2006), and Bauer (1993). Based on answers to the frequency of exposure and time spent on each individual ensemble-based fundamental exercise and on ensemble-based fundamentals as a whole, the relationship between ensemble-based fundamentals and performance achievement outcomes will be more clear. The EFS was sent only to high school band directors ($N \approx 34$) via Google Forms using convenience sampling. Requests for participation were sent via email, Facebook, Twitter, and through word of mouth, and the first 34 respondents were selected for participation. High school band directors were ideal participants because they are more likely to engage their bands in official contests/assessments (Austin, 1990; Banister, 1992; Saidon & Shah, 2013; Sheldon, 1994). Contests/assessment results are also more likely to be a contributing factor in the assessment of teaching quality for high school band directors (Austin, 1990; Saidon & Shah, 2013). Further, high school band contests/assessments often provide high quality recordings of performances. Each high school band director completing the EFS was asked to send a recording of a representative performance from a contest/assessment. These recordings played an important role in linking the relationship between the ensemble-based fundamentals practiced by each band to their ensemble performance achievement. The responses of each high school band director were paired with performance achievement ratings drawn from the Band Rating Rubric in order to explore the relationship between exposure to ensemble-based

fundamentals and ensemble performance achievement. The Ensemble Fundamentals Survery can be found in Appendix J.

Band Rating Rubric

Each participating high school band director was asked to submit a representative recording of their ensemble following their completion of the Ensemble Fundamental Survey. This process allowed the recordings, once rated, to be linked back to what was actually happening in the classroom. This link allowed the relationship between exposure to ensemble-based fundamentals and ensemble performance achievement to be explored more thoroughly.

In order to rate recordings that were submitted, a rating tool was needed. The Band Rating Rubric, developed by Edwards (2017) was used. This scale was used because it displays overall good psychometric qualities. In this scale, the reliability of the measures was addressed by the information functions of the parameters used to estimate the value of the latent trait, as evidenced by high reliability of separation for persons and items. (Edwards, 2017). Precision evidence was strong, in that only small standard errors were associated with each person and item (Edwards, 2017). When combined, the reliability and precision evidence indicate that the measure strongly separates performances based on an ordered hierarchy along the latent continuum using an equal-interval, logit gradation (i.e., linear measure; Edwards, 2017; Wesolowski et al., 2017). This ordering is an important component of predictive validity as the ordering of items according to difficulty can now be predicted prior to data collection with sample independence (Edwards, 2017). Reliability, precision evidence, and predictive validity is common when using this type of Rasch model process (Wesolowski et al., 2017). The Band Rating Rubric was used by a cohort of content area experts (N=36) to rate each recording submitted by participating high school band directors. Each content area expert was highly

qualified through years of experience (minimum of 8), previous adjudication experience, and prior adjudication training. Content area experts were each responsible for rating four recordings using the Band Rating Rubric. Recordings were interlocked (i.e., rater 1 will rate recordings 1-4, rater 2 will rate recordings 3-6, rater 3 will rate recordings 5-8, etc, with the final rater looping back to rate recordings 34, 35, 1, and 2) so that each recording was rated by at least two raters. Once each recording was rated, the ratings were converted to linear measure using the Many-Facet Rasch Partial Credit Model. The Band Rating Rubric can be found in Appendix K.

Data Analysis

The use of the Many-Facet Rasch (MFR) model was decided upon because, when compared to Classical Test Theory, it offers a theory that is more grounded (Wesolowski et al., 2016b). Rasch measurement theory is also preferred in scale development and in the measurement of latent traits in the behavioral, social, and health sciences, making it an ideal choice (Engelhard, 2013). Analysis of the rating data was performed by the computer program *FACETS* (Linacre, 2014). The primary benefit of using the Rasch model is that invariant measurement is achieved when adequate fit to the model is observed. In terms of assessments, invariant measurement means that particular items do not influence the measurement of participants who happen to take them, and that raters do not have an influence on the items when they are using them to measure (Wesolowski et al., 2016b). The property of invariant measurement that makes the Rasch model so valuable must be evaluated in empirical data. Invariant measurement requires a hypothesis that must be confirmed or disconfirmed by evidence in a data set (Engelhard, 1994). A set of five requirements that can be used to determine the degree to which invariant measurement is obtained for persons and items was developed by Engelhard and Perkins (2011). These requirements are (a) item-invariant measurement of

persons (i.e., the participants being measured must be independent of the items used for the measurement); (b) non-crossing person response functions (i.e., a participant who is more able must have a better chance of success than a participant who is less able); (c) person-invariant calibration of test times (i.e., item calibration must be independent of participants used for calibration); (d) non-crossing item response functions (all participants must have a better chance of success on an easy item than a difficult item); and (e) variable map (i.e., participants and items must be located simultaneously on a single underlying latent variable). Taking these five requirements into account shows that the benefit of Rasch approaches to measurement is the strong requirement that items being used can measure a single latent trait, the independence of items, and sample-independent estimations of person and item parameters (i.e., invariant measurement) (Wesolowski et al., 2016b).

Because raters were used in the assessment process, the Many-Facet Rasch (MFR) model can be used to define person ability, item difficulty, and rater severity simultaneously (Linacre, 1989/1994). For polytomous items, such as the rating scale used in this study, rating scale formulation of the model specifies consistent distances between rating scale categories across items (Wright & Masters, 1982). Once collected, data can be placed along a linear scale, correlations and regressions can be run using true scores in order to examine the relationship between exposure to ensemble-based fundamentals and ensemble performance achievement.

Once gathered and converted to linear measure using Rasch, data taken from the Ensemble Fundamentals Survey and the Band Rating Rubric was analyzed in order to answer the following research questions: (a) Is there an overall relationship between exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement?; (b) What is the relationship between exposure to (e.g., time and type) varying types of ensemble-based

fundamentals and ensemble performance achievement?; and (c) How well can performance achievement be predicted based upon exposure to (e.g., time and type) ensemble-based fundamentals.

To answer the first question, responses from the EFS detailing whether or not a high school band director engages their ensemble in ensemble-based fundamentals were correlated with that band's overall performance achievement score. Spearman's rho was used because it is more accurate when including categorical variables such as those taken from the EFS. This statistic helped to determine the direction and magnitude of a potential relationship between the practice of ensemble-based fundamentals in general and ensemble performance achievement. This relationship is important to explore as it helps determine if the practice or omission of ensemble-based fundamentals has a positive or negative relationship with performance achievement. For those high school band directors indicating that they do engage in ensemble-based fundamentals, the relationship between frequency (e.g., how often ensemble-based fundamentals are practiced) and performance achievement can also be explored, as can the relationship between time spent (e.g., how many minutes ensemble-based fundamentals are practiced during rehearsals in which they are used) and performance achievement. The direction and magnitude of these correlations can help explore, in an overarching way, whether or not practicing ensemble-based fundamentals in general tends to impact performance achievement in a positive or negative way.

The second research question was answered by exploring each individual ensemble-based fundamental, its frequency of use, and its time of exposure when used. Each fundamental that is used by a high school band director was correlated, again using Spearmans' rho, with that band's performance achievement score. Exploring these relationships helped to discover if there are

certain types of fundamental exercises that have a more significant positive or negative relationship with performance achievement. The direction and magnitude of any relationship may help discover if specific types of ensemble-based fundamental exercises are more valuable to engage in because they have a more positive relationship with ensemble performance achievement.

Answering the first two research questions helps to answer the third. Using statistically significant results from the first two results, an optimal scaling regression was used to determine if there was a predictive relationship between exposure to (e.g. time and type) applied ensemble fundamentals and ensemble performance achievement. For this test, each statistically significant applied ensemble fundamental type was treated as a separate independent variable, and performance achievement scores were the dependent variable. The optimal scaling regression was used because, as part of its process, it transforms categorical variables (e.g., responses to EFS) to linear data, and is thus more accurate than multiple regression for the purposes of this study.

CHAPTER 4

RESULTS

The purpose of this study was to explore the relationship between applied classroom ensemble fundamentals and ensemble performance achievement of high school bands. This study was guided by the following research questions: (a) is there an overall relationship between exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement?; (b) what is the relationship between exposure to (e.g., time and type) varying types of ensemble-based fundamentals and ensemble performance achievement?; and (c) how well can performance achievement be predicted based upon exposure to (e.g., time and type) ensemble-based fundamentals. Data was drawn from two tools: the Ensemble Fundamentals Survey (EFS) found in Appendix J, and the Band Rating Rubric found in Appendix K.

Descriptive Statistics

To answer the research questions, responses from participating band directors ($N = 34$) to question on the Ensemble Fundamentals Survey (EFS) were compiled and coded. Participating band directors responded to questions about the frequency (i.e., how many times per week) with which they practiced twelve types of ensemble-based fundamentals and how much time they spent practicing them when they were practiced. Recordings of each participating band director's band were rated by subject matter experts ($N = 36$) using the Band Rating Rubric. These ratings were analyzed using the Multi-Facet Rasch Model in order to transform categorical variables into a linear ensemble performance achievement measure. To answer the research questions, band director responses to the EFS were correlated with ensemble performance achievement to

determine if relationships existed. Statistically significant results were used in an optimal scaling regression to determine if a predictive relationship existed between any type of exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement. To answer the first research question, frequency (e.g., number of times per week) of exposure and time spent when exposed were considered separately. Box plots showing mean and quartile information for each of the twenty-four independent variables correlated with performance achievement can be found in Appendix O. To answer the second research question, frequency (e.g., number of times per week) of exposure and time spent when exposed were considered together. Box plots showing mean and quartile information for each of the twelve independent variables correlated with performance achievement can be found in Appendix P. To answer the third research question, statistically significant results from the first two research questions were used as independent variables in an optimal scaling regression with performance achievement used as the dependent variable. Box plots for the optimal scaling regressions showing mean and quartile information for each of the five independent variables can be found in Appendix Q.

Rasch Analysis Results

Multi-Facet Rasch (MFR) analysis using *FACETS* (Linacre, 2014) was used to transform observed categorical scores into a linear performance achievement score for each band. Categorical scores were taken from the responses of blind raters using the Band Rating Rubric. Table 4.1 presents summary statistics for the MFR output. Table 4.2 represents the calibration of rated recordings as determined by MFR analysis. Calibration of raters can be found in Appendix M. Calibration of Items can be found in Appendix N. Figure 4.1 is a variable map that graphically displays the latent variable investigated in this study.

Table 4.1

Summary Statistics for MRF Model

	Facets		
	Raters	**Recordings**	**Items**
Measure (Logits)			
Mean	0.00	-1.04	0.00
SD	1.40	1.58	0.76
N	36	34	23
Infit *MSE*			
Mean	1.00	1.00	0.99
SD	0.27	0.23	0.14
Std. Infit			
Mean	-0.20	-0.10	-0.10
SD	1.8	1.5	1.2
Outfit *MSE*			
Mean	1.06	1.06	1.06
SD	0.51	.046	0.37
Std. Outfit			
Mean	-0.20	-0.10	0.20
SD	1.7	1.4	1.6
Separation Statistics			
Reliability of Separation	0.97	0.98	0.95
Chi-Square	1205.8*	1846.7*	402.4*
Degrees of Freedom	35	33	22

*$p<0.01$

Table 4.2

Calibration of Rated Recordings

Recording Number	Observed Average	Measure	SE	Infit *MSE*	Std. Infit	Outfit *MSE*	Std. Outfit
15	1.50	3.42	0.22	0.83	-1.10	0.82	-0.90
13	1.8	2.15	0.20	1.19	1.30	1.19	1.20
21	1.4	2.04	0.22	0.99	0.00	0.87	-0.70
20	2.20	0.69	0.21	0.72	-2.10	0.69	-1.90
22	1.80	0.62	0.20	0.95	-0.30	0.92	-0.50
34	1.70	0.60	0.17	0.90	-0.90	0.90	-0.70
29	2.00	0.59	0.23	0.66	-2.30	0.71	-1.80
16	2.50	-0.28	0.22	0.99	0.00	1.07	0.40
27	2.00	-0.31	0.20	1.35	2.10	1.27	1.60
30	2.00	-0.42	0.18	0.79	-1.80	0.77	-1.90
32	2.20	-0.50	0.18	0.88	-0.90	0.89	-0.80
23	2.00	-0.59	0.20	0.74	-1.80	0.71	-1.90

Recording Number	Observed Average	Measure	SE	Infit *MSE*	Std. Infit	Outfit *MSE*	Std. Outfit
4	2.10	-0.63	0.20	0.89	-0.80	0.86	-1.00
26	2.10	-0.69	0.20	0.83	-1.10	0.79	-1.40
18	2.70	-1.03	0.26	1.18	1.00	1.12	0.40
6	2/10	-1.05	0.21	1.02	0.10	1.01	0.10
2	2.20	-1.18	0.20	0.73	-2.00	0.68	-2.30
17	2.60	-1.27	0.20	1.11	0.90	0.85	-0.40
25	2.30	-1.55	0.20	1.12	0.90	1.09	0.60
5	2.30	-1.65	0.21	1.17	1.00	1.15	0.80
24	2.4	-1.77	0.20	1.37	2.40	1.38	2.40
8	2.20	-1.77	0.20	1.17	1.10	1.19	1.20
9	2.30	-1.90	0.20	0.93	-0.40	0.94	-0.30
11	2.60	-1.97	-0.23	1.20	1.30	1.15	0.60
31	2.50	-2.00	0.22	0.62	-2.90	0.59	-2.20
14	2.90	-2.23	0.29	1.68	3.10	1.96	1.90
33	2.50	-2.44	0.17	1.07	0.60	1.03	0.20
19	2.80	-2.48	0.29	0.72	-1.60	3.14	2.10
3	2.50	-2.49	0.23	0.64	-2.50	0.53	-2.20
12	2.70	-2.60	0.25	1.06	0.40	0.76	-0.60
28	2.70	-2.75	0.22	1.30	2.10	1.43	2.00
1	2.50	-2.98	0.19	1.16	1.20	1.52	2.60
7	2.60	-3.36	0.22	1.05	0.40	1.06	0.30
10	2.70	-3.43	0.23	0.93	-0.40	0.97	0.00
Mean	2.30	0.00	0.22	1.00	-0.20	1.06	-0.20
SD	0.30	1.40	0.03	0.27	1.80	-.51	1.70

Note. The items are presented in Measure order, from high (highest achievement) to low (lowest achievement).

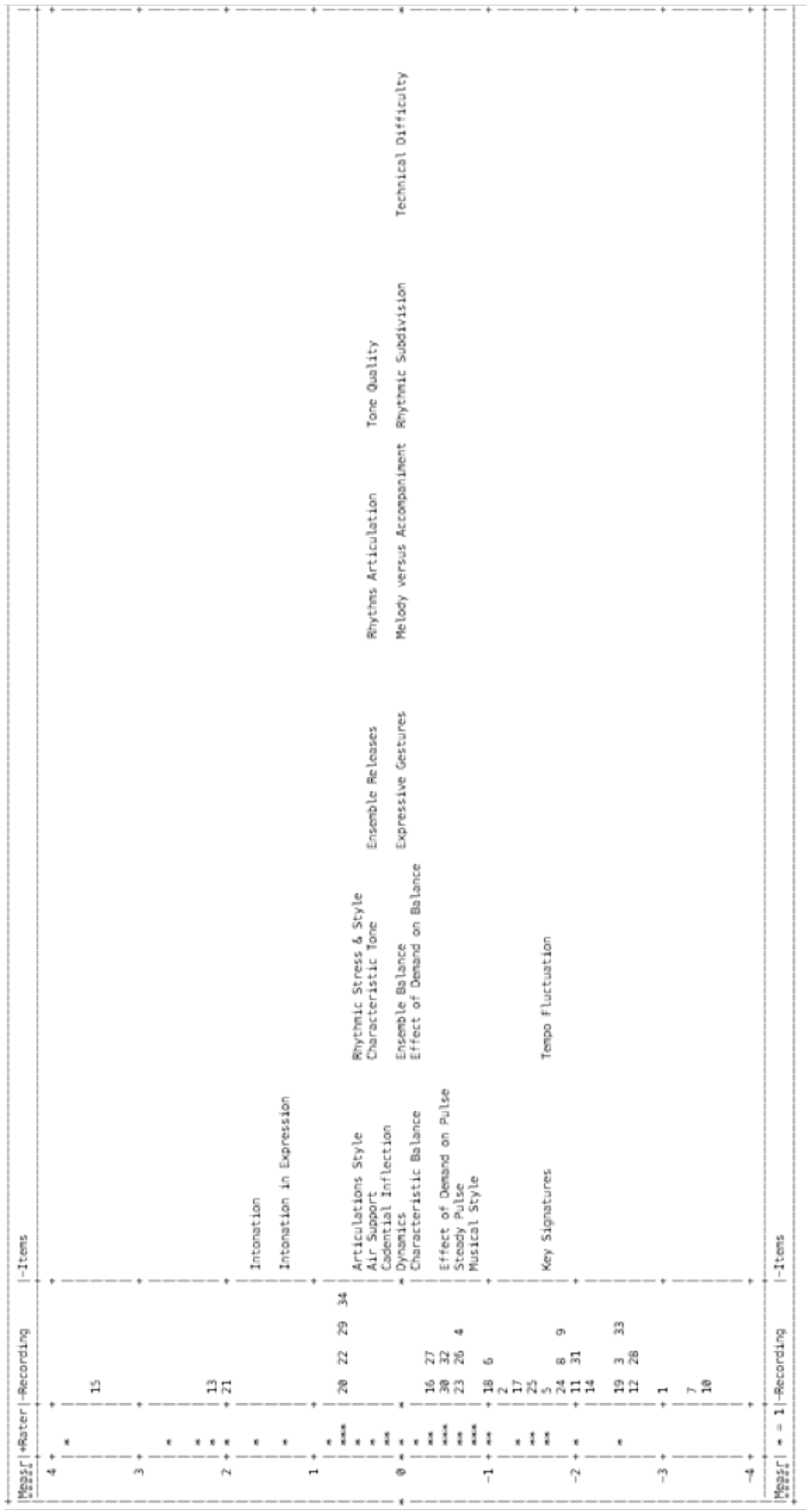


Figure 4.1 Variable Map

Research Question 1

The first research question was: is there an overall relationship between exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement? To answer this question, coded EFS data for each band was correlated with the bands' performance achievement score using Spearman's rho to help account for the use of categorical variables. Each participating band director ($N = 34$) responded to the frequency (i.e., how many times per week) with which they practice each of 12 ensemble-based fundamentals, as well as the time (i.e., how many minutes) spent on each fundamental when it was practiced. For this question, frequency (i.e., how many times per week) of exposure and time spent (i.e., number of minutes when exposed) were considered separately in order to determine if there was an overall relationship based on any exposure type. Table 4.3 shows that there was a statistically significant relationship between the frequency of practice of balance and blend exercises and ensemble performance achievement ($p < .01$), the time spent on scales and arpeggios and ensemble performance achievement ($p < .05$), and the time spent on tuning exercises and ensemble performance achievement ($p < .05$). The information in table 4.3 applies directly to the first research question.

In this case, the direction of the each statistically significant correlation was negative, meaning that an inverse relationship existed between each applied ensemble fundamental listed and performance achievement score. This assumes that the ensemble performance achievement score will decrease with a corresponding increase in frequency of exposure to balance and blend exercises, and/or an increase in time spent practicing scales and arpeggios, and/or an increase in time spent practicing tuning exercises. These results were contrary to what was expected, and possible reasons for that contradiction are discussed in Chapter 5.

Table 4.3

Spearman's Rho Correlation Between Ensemble Fundamental Type and Performance Achievement Score

Ensemble Fundamental Type	Performance Achievement Score
Freq – Long tones	-0.01
Freq – Lip Slurs	0.07
Freq – Scales/Arpeggios	0.08
Freq – Tuning	-0.28
Freq – Chorales	0.20
Freq – Sight-Reading	0.02
Freq – Rhythm Drills	0.07
Freq – Articulation	0.23
Freq – Breathing	0.10
Freq – Dynamics	0.19
Freq – Balance/Blend	**-0.56
Freq – Singing	-0.04
Time – Long Tones	<-0.01
Time – Lip Slurs	-0.04
Time – Scales/Arpeggios	*-0.36
Time – Tuning	*-0.42
Time – Chorales	0.08
Time – Sight-Reading	-0.02
Time – Rhythm Drills	-0.12
Time – Articulation	-0.09
Time – Breathing	-0.13
Time – Dynamics	-0.02
Time – Balance/Blend	-0.18
Time – Singing	-0.26

Note. Frequency refers to how often fundamental was used. Time refers to how long fundamental was practiced when it was used. Shaded rows show statistically significant results.
** $p < 0.01$, * $p < 0.05$

Research Question 2

The second research question was: what is the relationship between exposure to (e.g., time and type) varying types of ensemble-based fundamentals and ensemble performance achievement? To answer this question, coded EFS data was again correlated with each bands'

performance achievement score using Spearman's rho. Each participating band director ($N = 34$) responded to the frequency (i.e., how many times per week) with which they practice each of 12 ensemble-based fundamentals, as well as the time (i.e., how many minutes) spent on each fundamental when it was practiced. For this question, data was coded to be inclusive of frequency (i.e., how many times per week) of exposure and time spent (i.e., number of minutes when exposed). Table 4.4 shows that there was a statistically significant relationship between tuning exercises and performance achievement ($p < .01$), and balance and blend exercises and performance achievement ($p < .01$). The information in Table 4.4 applies directly to the second research question.

Table 4.4

Spearman's Rho Correlation Between Ensemble Fundamental Type and Performance Achievement Score

Ensemble Fundamental Type	**Performance Achievement Score**
Long Tones	0.01
Lip Slurs	0.07
Scales/Arpeggios	0.07
Tuning Exercises	**-0.46
Chorales	0.24
Sight-Reading	0.01
Rhythm Drills	0.06
Articulation	0.16
Breathing Exercises	<0.01
Dynamics	0.16
Balance/Blend	**-0.46
Singing	-0.08

Note. Ensemble Fundamental Type coded to be inclusive of frequency and time spent. Shaded rows show statistically significant results.
** $p < 0.01$

In this case, the direction of each statistically significant correlation was negative, meaning that an inverse relationship existed between the time and type of exposure to each

applied ensemble fundamental listed and performance achievement score. This assumes that performance achievement score will decrease with a corresponding increase in exposure to tuning exercises and/or balance and blend exercises. These results were also contrary to what was expected. Possible reasons for the contradictions are discussed in Chapter 5.

Research Question 3

The third research question was: how well can performance achievement be predicted based upon exposure to (e.g., time and type) ensemble-based fundamentals? Each participating band director ($N = 34$) responded to the frequency (i.e., how many times per week) with which they practice each of 12 ensemble-based fundamentals, as well as the time (i.e., how many minutes) spent on each fundamental when it was practiced. To answer this question, statistically significant relationships from the first two research questions were run through an optimal scaling regression to determine predictive validity. Optimal scaling regression was chosen because as part of its process it transforms categorical variables (such as those taken from the EFS) into linear measure. For this question, frequency (i.e., how many times per week) of exposure and time spend (i.e., number of minutes when exposed) were considered both together and separately to help determine if any type of exposure to ensemble-based fundamentals had a predictive relationship with performance achievement. Frequency of exposure to balance and blend, time spent with scales and arpeggios, and time spent with tuning exercises, overall exposure to tuning exercises, and overall exposure to balance and blend were treated as independent variables and performance achievement scores were treated as the dependent variable. Table 4.5 shows the standardized coefficients for statistically significant relationships when time and type were considered separately.

Table 4.5

Summary of Optimal Scaling Regression Analysis for Variables Predicting Performance Achievement

	Performance Achievement Score				
Variable	*B*	*SE B*	df	F	*p*
Freq – Balance/Blend	-0.51	0.18	2	7.33	< 0.01*
Time – Scales/Arpeggios	-0.18	0.25	1	0.55	0.47
Time – Tuning Exercises	-0.20	0.37	1	0.27	0.61

Note. Frequency refers to how often fundamental was used. Time refers to how long fundamental was practiced when it was used. Shaded row shows statistically significant result.
*$p < 0.01$

When considered separately for time and type of exposure, frequency of exposure to balance and blend exercises had a highly significant ($p < .01$) negative ($B = -.509$) predictive relationship with performance achievement. This means that as frequency of exposure to balance and blend exercises increases, performance achievement will correspondingly decrease. This was contrary to what was expected, and that contradiction will be discussed in Chapter 5.

Table 4.6 shows the summary of the optimal scaling regression analysis when data was coded to be inclusive of frequency and time spent. When considering frequency and time of exposure together, overall exposure (i.e., frequency (i.e., how many times per week) of exposure and time spent (i.e., number of minutes when exposed)) to balance and blend exercises had a highly significant ($p < .01$) negative ($B = -.974$) predictive relationship with performance achievement. This means that overall exposure to balance and blend exercises could predict a corresponding reduction in performance achievement score according to this model. This was contrary to what was expected, and possible reasons for that contradiction are discussed in Chapter 5.

Table 4.6

Summary of Optimal Scaling Regression Analysis for Variables Predicting Performance Achievement

	Performance Achievement Score				
Variable	*B*	*SE B*	df	F	*p*
Tuning Exercises	0.55	0.45	2	1.50	0.24
Balance/Blend	-0.97	0.32	4	9.15	<0.01*

Note. Variables coded to be inclusive of frequency and time spent. Shaded row shows statistically significant result.
*$p < 0.01$

Overall, the results of the optimal scaling regression both showed a highly significant ($p < .01$) negative predictive relationship between exposure to balance and blend exercises and ensemble performance achievement. Therefore, according to the results of this study, exposure to balance and blend exercises can help predict performance achievement scores.

CHAPTER 5

DISCUSSION

Introduction

The purpose of this study was to explore the relationship between applied classroom ensemble fundamentals and ensemble performance achievement of high school bands. Data was collected via the Ensemble Fundamentals Survey (EFS), a tool with which high school band directors were able to describe exactly what types of ensemble-based fundamental exercises they use, how often they use them, and for how long when they are used. Participating band directors also sent recordings from official contests/assessments which were then blindly rated by content area experts using the Band Rating Rubric (BRR). Observed scores from the BRR were transformed to linear measure through application of the Multi-Facet Rasch Model (Rasch, 1960/1980) in order to determine linear performance achievement scores. Spearman's rho correlations were run to determine significance of relationships, and optimal scaling regression was used to determine if any significant relationships were predictive. The results showed that there was an overarching relationship between applied ensemble-based fundamentals and performance achievement, that significant relationships were negative in direction, and that at least one ensemble-based fundamental (balance and blend exercises) had a predictive relationship with ensemble performance achievement. First, this study discussed existing literature pertaining to applied ensemble-based fundamentals, how they are (or are not) used in the band classroom, and arguments for and against their use. Also discussed was extant literature about various types of competitive and non-competitive band contests, as well as literature

pertaining to expert raters who adjudicate such contests and who were used as blind raters for the purposes of this study. Next, an overview of the methodology was provided which included information regarding subjects, the tools used for data collection, the process of data collection, the coding of data, and data transformation using MFR. Results of the data analysis were provided in Chapter 4. This chapter will summarize the study and present findings to the three research questions, identify possible limitations of the study, and discuss suggestions for future research in the area of the relationship of applied ensemble-based fundamentals and ensemble performance achievement.

Research Questions

Research Question 1. Is there an overall relationship between exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement? A Spearman's rho correlation between all different types of ensemble-based fundamentals and their frequency and time of exposure determined that there is an overall relationship between exposure to (e.g., time and type) ensemble-based fundamentals and ensemble performance achievement. It was expected that there would be several ensemble-based fundamental exercise types that would correlate highly in a positive direction with ensemble performance achievement. The results unexpectedly showed that only three types of ensemble-based fundamentals (frequency of exposure to balance and blend exercises, time spent with scales and arpeggios, and time spent with tuning exercises) had a relationship with performance achievement, and each of them was negative in direction. This means that the model showed that as frequency or time of exposure to those three ensemble fundamentals increased, ensemble performance achievement decreased.

The results were particularly surprising because much previous research and particularly that of Love (1997) suggests that engaging in ensemble-based fundamentals was important for

higher quality ensemble performance achievement. While all ensemble-based fundamentals included in the EFS for this study were drawn from previous research, it is possible that some of the outcomes that each individual ensemble-based fundamental provided in the tool could be nested within other ensemble-based fundamentals included in the survey. This potential nesting could cause some ensemble-based fundamentals to be represented inaccurately in survey results and therefore confound results of the Spearman's rho correlation.

Research Question 2. What is the relationship between exposure to (e.g., time and type) varying types of ensemble-based fundamentals and ensemble performance achievement? To answer this question, responses from the EFS were coded to be inclusive of the frequency (e.g., how many times per week) with which a particular ensemble-based fundamental was practiced and how much time was spent practicing it when it was practiced. A Spearman's rho correlation was run to determine what type of relationship, if any, existed between exposure to (e.g., time and time) varying types of ensemble-based fundamentals and ensemble performance achievement. It was expected that there would be a positive relationship between exposure to ensemble-based fundamentals and ensemble performance achievement in such a way that more exposure to certain types of ensemble-based fundamentals would correlate with higher quality ensemble performance achievement. Surprisingly, the opposite result was found. Two ensemble-based fundamentals had statistically significant relationships (balance and blend exercises and tuning exercises) with ensemble performance achievement, but in both cases the direction of the relationship was negative. The results indicated that more exposure to these two ensemble-based fundamentals would correspond to a decrease in ensemble performance achievement.

Much like the results of research question 1, these results were surprising because previous research suggested the opposite outcome was more likely. The potential nesting of

ensemble-based fundamentals as mentioned in the discussion for research question 1 could help to account for the unexpected result.

Research Question 3. How well can performance achievement be predicted based upon exposure to (e.g., time and type) ensemble-based fundamentals? To answer this question, the five total significant relationships derived from the first two research questions were used in an optimal scaling regression which transformed the categorical variables drawn from the EFS into independent linear measures that were run against the dependent variable of performance achievement. The results showed that balance and blend exercises had a highly significant predictive relationship with ensemble performance achievement. It was expected that this relationship would be positive, but in fact it was negative, and with high magnitude. This means that exposure to balance and blend exercises was able to predict performance achievement. As that exposure increases, ensemble performance achievement should decrease.

The result here was again surprising because previous research indicated that exposure to ensemble-based fundamentals had a positive relationship with ensemble performance achievement. The idea of nesting variables was particularly important here, because the concept of balance and blend could be defined in multiple ways and practiced in multiple ways. For example, an ensemble could practice balance and blend while doing chorale exercises, tuning exercises, or nearly any of the other ensemble-based fundamentals included in the EFS. The result of the optimal scaling regression, then, might indicate a negative predictive relationship between exposure to balance and blend exercises and ensemble-performance achievement because those band directors that indicated that they spent more time specifically on balance and blend as part of their fundamental exercise time in rehearsal achieved at lower levels than those bands that nested balance and blend studies into other ensemble-based fundamental exercises.

Implications of the Findings

This study showed that applied ensemble-based fundamentals have a relationship with ensemble performance achievement, that the only statistically significant relationships were negative in direction, and that at least one ensemble-based fundamental (balance and blend exercises) has a negative predictive relationship with performance achievement. Based on these results, it would be easy to assume that the value of practicing ensemble-based fundamentals on a regular basis has been overstated when it comes to their relationship with performance achievement. The results of the data in this study seem to back up that assumption. But to make that assumption would carry with it another assumption that the specific way in which ensemble-based fundamentals are practiced is the same in every rehearsal hall and with every band director. However, the work of Love (1997), Garofalo & Whaley (1979), Manfredo (2006), Lane (2010) and others indicate that there exists a diversity in the implementation, structure, pacing, and motivation behind the use of ensemble-based fundamentals in the high school band classroom. It is unlikely, then, that the argument that all ensemble-based fundamentals are practiced the same way can be made and backed up with fact. Additionally, as the variable map shows, many of the latent item traits are found above the current level of individual ensembles performance achievement, thereby contributing to the surprising results.

The results of this study more likely imply that it does not matter which ensemble-based fundamentals are practiced, how often they are practiced, and for how long they are practiced in those instances. What does matter is how the ensemble-based fundamentals that are used are practiced. Nesting multiple fundamental concepts into a single exercise (e.g., using a chorale exercise to practice balance and blend, intonation, and articulation or using scales and arpeggio exercises to also practice articulation, rhythm, and breathing) not only gives a band director the

ability to work on more individual ensemble-based fundamentals in a shorter amount of time, it helps them to promote a consistently high level of expectation and demands which can lead to higher quality ensemble performance achievement (Yarbrough & Price, 1989). Therefore, more emphasis should be placed on how ensemble-based fundamentals are practiced efficiently and effectively. The results of this study, and particularly the negative predictive relationship between balance and blend exercises and performance achievement, seem to back up that practicing ensemble-based fundamentals without embedded-musicianship by way of nested fundamental concepts simply takes time out of rehearsal and results in lower quality ensemble-performance achievement.

Limitations

While the results of this study were statistically strong, they were unexpected. That could be due to several limitations of the study. The first limitation was within the design of the EFS itself. The EFS was designed to be short and concise, so that busy band directors would be willing to take the time to participate in the study. Although previous research literature was used to develop the list of ensemble-based fundamentals, frequency periods, and time exposure breakdowns included in the EFS, the list was neither exhaustive nor did it take into account potential nesting of ensemble-based fundamentals. The EFS also did not try to determine how ensemble-based fundamentals were taught and practiced, largely because that would have made the survey much longer and likely reduced subscription. Lastly, the EFS did not ask at what point in the pacing of a normal rehearsal band directors engaged their bands in ensemble-based fundamentals. As the variable map (Figure 4.1, page 36) shows, many of the latent item traits are found above the current level of individual ensembles performance achievement. Therefore, band directors practicing those items could actually be hurting their overall ensemble

performance achievement by spending time on tasks that are too far above the level of the ensemble. The sample size of participating band directors ($N = 34$) was on the small side, and although the thirty-four blindly rated recordings were enough to run a reliable and valid MFR analysis, more participation on the part of band directors may have helped identify more (and potentially more positive) trends in the use of ensemble-based fundamentals in the high school band classroom. Finally, the study was narrowed to only high school bands and band directors, and could have been expanded to include more age and ability groups.

Suggestions for Future Research

This study is the first or one of the first to directly tie what high school band directors say they do in the classroom to a linear performance achievement score based on that specific bands' performance. First, a study such as this could be expanded to include more age and ability levels in order to dig into potential trends regarding what happens in the band classroom and the same band's performance achievement scores. It would be wise to continue with studies that seek to tie what happens in the classroom regarding ensemble-based fundamentals to linear performance achievement outcomes. In order to more thoroughly explore what happens in the classroom, a new survey should be designed to help account for more variance in how ensemble-based fundamentals are used in the classroom, as well as to account for what other activities take place during a routine day in a given band classroom. It could be discovered that the activities outside of the time allotted for ensemble-based fundamental practice have a more significant positive relationship with performance achievement. To more reliably investigate the relationship between ensemble-based fundamentals and performance achievement, future research should dig more deeply into exactly how high school band directors teach, practice, and use ensemble-based fundamentals. The most efficient way to achieve this would be to obtain and code videos from a

large sample of band rehearsals which feature a wide variety of time spent on ensemble-based fundamentals and on types of ensemble-based fundamentals practiced. The researcher could then codify exactly how ensemble-based fundamentals are being taught and practiced, and compare that knowledge to ensemble performance achievement scores. A deeper understanding of how ensemble-based fundamentals are used would likely draw stronger (and potentially more positive) relationships with performance achievement. An effort should also be made to drill down into the nesting of ensemble-based fundamentals into one another in order to discover which individual fundamental exercises are most often nested in others and to determine if there are trends in this area. This could be done through a more specific survey tool or through the collection and coding of rehearsal videos. Understanding how individual ensemble-based fundamentals are nested into one another could further clarify their relationship with performance achievement. Further, an examination of the specific exercises used to teach and practice ensemble-based fundamentals, rather than a broad ensemble-based fundamental category, might also uncover a relationship with ensemble performance achievement. Most importantly, all data analyzed should be linear or transformed in to linear measure using the Many-Facet Rasch Measurement Model in order to produce reliable and valid results. There is still much work to be done in this area; however, this study could be an important first step to directly tying what happens in the classroom to performance achievement outcomes and eventually to contest results. In this way, the study of the use of ensemble-based fundamentals and their relationship with ensemble performance achievement can have a positive and lasting impact on band pedagogy and curriculum development

Conclusion

There have been previous studies conducted that research what types of ensemble-based fundamentals are used in the high school band classroom (Bauer, 1993; Blocher et al., 1997; Love, 1997). There have also been previous studies conducted that seek to correlate band classroom teaching with ensemble performance achievement through contest or assessment scores (Banister, 1992; Price & Chang, 2005b; Saidon & Shah, 2013). However, this was the first or one of the first studies to attempt to directly tie the ensemble-based fundamental exercises that high school band directors report using to a linear ensemble performance achievement score. In order to determine a linear performance achievement scores, performances from the bands of participating band directors were blindly rated by expert raters using the Band Rating Rubric (BRR). Raw data from the BRR was transformed into a linear measure using the Multi-Facet Rasch (MFR) measurement model (Rasch, 1960/1980). This study sought to explore the relationship between the use of applied ensemble-based fundamentals and performance achievement.

The strength and validity of the use of the MFR model to transform categorical data into linear measure should encourage educational researchers to consider using Rasch to help make categorical data more meaningful. Currently, categorical data is often treated as linear measure which leads to less than reliable results. The results of this MFR analysis of responses from blind raters to recordings using the BRR provided a true linear ensemble performance achievement measure. This linear measure allowed for direct correlation with responses to the Ensemble Fundamentals Survey without the fear of rater bias or “ratings creep”. Statistically significant results could then be reliably put through optimal scaling regression to determine predictive relationships. The results of these statistical analyses showed negative correlations between

certain ensemble-based fundamentals and ensemble performance achievement. Further, the specific ensemble-based fundamental "balance and blend exercises" had a negative predictive relationship with ensemble performance achievement, meaning that more exposure to balance and blend exercises would predict lower performance achievement outcomes. The use of MFR and optimal scaling regression helped to increase reliability by transforming categorical data into more meaningful linear measures, increasing confidence in the results of the study.

While the results of the study were surprising, they may point to the fact that how applied-ensemble based fundamentals are taught and practiced is more important than the specific exercises that are used in practice. Individual band directors may have unique ways of nesting ensemble-based fundamentals within one another, and the process of nesting likely confounded the results of this study as the survey tool did not account for it. This nesting likely allows the band directors to more efficiently cover a wider range of ensemble-based fundamentals without using an out-of-proportion amount of rehearsal time to do so. In this situation, band directors would be able to efficiently practice a variety of ensemble-based fundamentals in a short amount of time and still be able to spend a significant portion of rehearsal time rehearsing literature being studied for performance. In a way, this is the best of both the views of the proponents and detractors of the consistent use of ensemble-based fundamentals in the band classroom. Nesting not only accounts for the desire to practice ensemble-based fundamentals every day, but leaves enough rehearsal time to thoroughly include literature rehearsal on a daily basis. Learning more about the nesting process and its potential impact on performance achievement is a fertile ground for future research in this area.

For too long, high school band directors have engaged in the practice of ensemble-based fundamental exercises without truly knowing their relationship with ensemble performance

achievement. This lack of knowledge can manifest itself in band directors using ensemble-based fundamentals simply because they were told to, had done so when they were in high school or college, or even because a successful peer teacher used them. This is often done with no evidence that the ensemble-based fundamentals being used are actually helping the band gain higher quality performance achievement. As the results of this study show, spending more time on ensemble-based fundamentals does not necessarily mean that performance achievement will improve, and therefore high school band directors might be engaging in a counterproductive practice when they spend more time on ensemble-based fundamentals without paying close attention to how they practice them. For equally as long as band directors have practiced ensemble-based fundamentals for the sake of practicing them, they have based the effectiveness of the use of those ensemble-based fundamentals on performance achievement scores that were categorical rather than linear, and therefore did not tell the whole story of the value of the relationship between ensemble-based fundamentals and ensemble performance achievement. While categorical scoring remains the norm in the band world, it would be worth exploring whether or not a system could be developed that both provided a linear score (or ranked bands along a linear scale) and gave band directors information that could be used to positively impact curriculum.

This study used a sound and objective design coupled with the use of specific linear statistics in order to attempt to identify specific types of ensemble-based fundamentals that have a positive predictive relationship with performance achievement. That outcome did not materialize. Instead, the unexpected results did identify a negative predictive relationship between exposure to balance and blend exercises and ensemble performance achievement. The unexpected results also helped to identify several specific areas in which future research can

build in order to better understand the relationship between applied ensemble-based fundamentals and ensemble performance achievement. Further objective research using linear measures rather than categorical data is needed. However, it may be that it is not which specific ensemble-based fundamentals are practiced, but how they are nested, practiced, and sequenced that is the true predictive link between ensemble-based fundamentals and ensemble performance achievement.

REFERENCES

Asmus, E. P. (1985). The development of a multidimensional instrument for the measurement of affective responses to music. *Psychology of Music*, *13*, 19–30.

Asmus, E. P. (1989). Factor analysis : A look at the technique through the data of rainbow. *Bulletin of the Council for Research in Music Education*, (101), 1–29.

Asmus, E. P. (1999). Music assessment concepts. *Music Educators Journal*, *86*(2), 19–24. http://doi.org/10.2307/3399585

Austin, J. R. (1990). Competition: Is music education the Lloser? *Music Educators Journal*, *76*(6), 21–25.

Banister, S. (1992). Attitudes of high school band directors toward the value of marching and concert band contests and selected aspects of the overall band program. *Missouri Journal of Research in Music Education*, *29*, 49–57.

Bauer, W. I. (1993). The Relationship between rehearsal procedures and contest ratings for high school bands. *Contributions to Music Education*, *20*, 32–44.

Bergee, M. J. (1992). A scale assessing music student teachers' rehearsal effectiveness. *Journal of Research in Music Education*, *40*(1), 5–13.

Bergee, M. J. (1995). Primary and higher-order factors in a scale assessing concert band performance. *Bulletin of the Council for Research in Music Education*, (126), 1–14.

Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, *56*(2), 137–150.

Bergee, M. J. (2007). Performer, rater, occasion, and sequence as sources of variability in music performance assessment. *Journal of Research in Music Education*, *55*(4), 344–358. http://doi.org/10.1177/0022429408317515

Bergee, M. J., & McWhirter, J. L. (2005). Selected influences on solo and small-ensemble festivalrRatings: Replication and extension. *Journal of Research in Music Education*, *53*(2), 177–190. http://doi.org/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*, 342–353.

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.

Blocher, L., Greenwood, R., & Shellahamer, B. (1997). Teaching behaviors of middle school and high school band directors in the rehearsal setting. *Journal of Research in Music Education*, *45*(3), 457–469.

Boeckman, J. (2002). Grade inflation in band contest ratings: A trend study. *Journal of Band Research*, *38*(1), 25–36.

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.

Brand, M. (1985). Development and validation of the home musical environment scale for use at the early elementary level. *Psychology of Music*, *13*(40–48).

Brendell, J. (1996). Time use, rehearsal activity, and student off-task behavior during the initial minutes of high school choral rehearsals. *Journal of Research in Music Education*, *44*(1), 6–14.

Brophy, J. E. (1979). Teacher behavior and its effects. *Journal of Educational Psychology*, *71*(6), 733–750. http://doi.org/10.1037/0022-0663.71.6.733

Burnsed, V., Hinkle, D., & King, S. (1985). Performance evaluation reliability at selected concert festivals. *Journal of Band Research*, *21*(1), 22–29.

Cassidy, J. W. (1990). Effect of intensity training on preservice teachers' instruction accuracy and delivery effectiveness. *Journal of Research in Music Education*, *38*(3), 164–174. http://doi.org/0038040710383521z

Cavitt, M. E. (2004). Information in rehearsal frames targting intonation performance. *Journal of Band Research*, *40*(1), 48–52.

Ciorba, C. R., & Smith, N. Y. (2009). Measurement of instrumental and vocal undergraduate performance juries using a multidimensional assessment rubric. *Journal of Intellectual Disabilities : JOID*, *57*(1), 5–15.

Conrad, D. (2003). Judging the judges: Improving rater reliability at music contests. *NFHS Music Association Journal*, *20*(2), 27–31.

DCamp, C. B. (1980). *An application of the facet-factorial appraoch to scale construction in the development of a rating scale for high school band performance*. (Doctoral dissertation, Ph. D. University of Iowa).

Duke, R. A., Prickett, C. A., & Jellison, J. A. (1998). Empirical description of the pace of music instruction. *Journal of Research in Music Education*, *46*(2), 265–280. http://doi.org/10.2307/3345628

Duke, R. A., & Simmons, A. L. (2006). The nature of expertise: Narrative descriptions of 19 common elements observed in the lessons of three renowned artist-teachers. *Bulletin of the Council for Research in Music Education*, *170*(170), 7–19.

Edwards, A. S. (2017). *The psychometric evaluation of a wind band performance rubric using the multifaceted rasch partial credit measurement model*. (Unpublished Doctoral dissertation, University of Georgia).

Edwards, J. S., & Edwards, M. C. (1971). A scale to measure attitudes towards music. *Journal of Research in Music Education*, *19*, 228–233.

Engelhard, G. (1994). Examining rater errors in the assessment of written composition with a many-faceted Rasch model. *Journal of Educational Measurement*, *31*(2), 93–112.

Engelhard, G. (2002). Monitoring raters in performance assessments. In G. Tindal & T. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (pp. 261–287). Mahwah, NJ: Erlbaum.

Engelhard, G. (2013). *Invariant measurement: Using Rash models in the social, behavioral, and health sciences*. New York NY: Routledge.

Engelhard Jr., G., & Perkins, A. F. (2011). Person response functions and the definition of units in the social sciences. *Measurement: Interdisciplinary Research & Perspective*, *9*, 40–45.

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*, *2*, 19–24.

Garofalo, R. J. (1973). Blueprint for band: We need sound curriculums, not just skilled performers, *Music Educators Journal, 60*(3), 38–42.

Garofalo, R. J., & Whaley, G. (1979). Comparison of the unit study and traditional approaches for teaching music through school band performance. *Journal of Research in Music Education*. http://doi.org/10.2307/3344964

Gillis, G. (2008). Conductor responsibilities and rehearsal preparation. *Canadian Music Educator*, *49*(4), 36–39.

Goolsby, T. W. (1996). Time use in instrumental rehearsals: A comparison of experienced, novice, and student teachers. *Journal of Research in Music Education*, *44*(4), 286–303.

Goolsby, T. W. (1999). A comparison of expert and novice music teachers' preparing identical band compositions: An operational replication. *Journal of Research in Music Education*, *47*(2), 174. http://doi.org/10.2307/3345722

Gutsch, K. U. (1965). Evaluation in instrumental music performance: An individual approach. *Bulletin of the Council for Research in Music Education*, (4), 21–29.

Hamann, D. L., & Banister, S. (1991). Factors contributing to high school band students' ratings at solo and ensemble festival. *Contributions to Music Education*, *18*, 57–65.

Hash, P. M. (2012). An analysis of the ratings and interrater reliability of high school band contests. *Journal of Research in Music Education*, *60*(1), 81–100. http://doi.org/10.1177/0022429411434932

Hash, P. M. (2013). Large-group contest ratings and music teacher evaluation: Issues and recommendations. *Arts Education Policy Review*, *114*(4), 163–169.

Iusca, D. (2014). The effect of evaluation strategy and music performance presentation format on score variability of music students ' performance assessment. *Procedia - Social and Behavioral Sciences*, *127*, 119–123. http://doi.org/10.1016/j.sbspro.2014.03.224

King, S., & 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), 27–33.

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*(4), 322–337.

Kruth, E. (1970). Adjudication and the music festival. *The Instrumentalist*, *24*(6), 48.

Lane, J. S. (2010). An analysis of relationships between lesson planning training and rehearsal pacing of undergraduate music education majors in practice teaching settings.

Latimer Jr., M. E., Bergee, M. J., & Cohen, M. L. (2010). Reliability and perceived pedagogical utility of a weighted music performance assessment rubric. *Journal of Research in Music Education*, *58*(2), 168–183.

Linacre, J. M. (1987). *Many-facet Rasch Measurement*. Chicago: MESA Press.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*, 85–106. http://doi.org/10.1037//1076-898X.8.2.99

Linacre, J. M. (2014). *Facets*. Chicago IL: MESA Press. http://doi.org/10.1016/S0022-3913(12)00047-9

Love, J. W. (1997). *The relationship between rehearsal structure and contest ratings for high school bands*. (Doctoral dissertation, University of Kansas).

Manfredo, J. (2006). Effective time management in ensemble rehearsals. *Music Educators Journal*, *93*(2), 42–46.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. http://doi.org/10.1007/BF02296272

Maynard, L. M. (2006). The role of repetition in the artist practice sessions of teachers and their students. *Bulletin of the Council for Research in Music Education*, *Winter*(167), 61–72.

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.

McPherson, G. E., & Hendricks, K. S. (2010). Students' motivation to study music: The United States of America. *Research Studies in Music Education*, *32*(2), 201–213.

McPherson, G. E., & Thompson, W. F. (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education*, 10(1), 12-24.

Meyers, B. D. (2011). *Attitudes of high school band directors in the United States toward solo and ensemble activities*. (Doctoral dissertation, Arizona State University).

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.

Nichols, J. P. (1991). A factor-analysis approach to the development of a rating scale for snare drum performance. *Dialogue in Instrumental Music Education*, *15*, 11–31.

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*(3), 237–251. http://doi.org/10.1177/002242940705500305

Opsal, S. K. (2013). *Competition and extrinsic motivation in the band classroom : A review of literature and suggestions for educational practice*. (Honors Program Theses, University of Northern Iowa).

Paschall, S. (2006). *A review of beginning band method books for inclusion of comprehensive musicianship and adherence to the national standards for music education*. (Doctoral dissertation, Bowling Green State University).

Price, H. E. (1983). The effect of conductor academic task presentation, conductor reinforcement, and ensemble practice on performers' musical achievement, attentiveness, and attitude. *Journal of Research in Music Education*, *31*(4), 245–257.

Price, H. E., & Chang, E. C. (2001). Conductor expressivity and ensemble performance: An exploratory study. *Contributions to Music Education*, *28*(2), 9–20.

Price, H. E., & Chang, E. C. (2005a). Conductor and ensemble performance expressivity and state festival ratings. *Journal of Research in Music Education*, *53*(1), 66–77.

Rasch. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institude for Educational Research (Expanded edition, 1980 Chicago, IL: University of Chicago Press).

Saidon, Z. L., & Shah, S. M. (2013). Assessment in competition as a tool for developing high school marching bands. *Procedia - Social and Behavioral Sciences*, *93*, 1005–1009.

Sandstrom, G. M., & Russo, F. A. (2013). Absorption in music: Development of a scale to identify individuals with strong emotional responses to music. *Psychology of Music*, *41*, 216–228.

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.

Shaw, C. N., & Tomcala, M. (1976). Music attitude scale for use with upper elementary school children. *Journal of Research in Music Education*, *24*, 73–80.

Sheldon, D. A. (1994). Effects of competitive versus noncompetitive performance goals on music students ' ratings of band performances. *Bulletin of the Council for Research in Music Education*, (121), 29–41.

Silveira, J. M. (2014). The perception of pacing in a music appreciation class and its relationship to teacher effectiveness and teacher intensity. *Journal of Research in Music Education*, *62*(3), 302–318. http://doi.org/10.1177/0022429414542978

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.

Smith, D. T. (2009). Development and validation of a rating scale for wind jazz improvisation performance. *Journal of Research in Music Education*, *57*, 217–235.

Teachout, D. J. (2001). The relationship between personality and the teaching effectiveness of music student teachers. *Psychology of Music*, (29), 179–192.

Umberson, G. (1970). Experimentalism and pacing in the choral rehearsal. *The Choral Journal*, *10*(6), 10–14.

Wesolowski, B. C. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, *98*(3), 36–42. http://doi.org/10.1177/0027432111432524

Wesolowski, B. C., Amend, R. M., Barnstead, T. S., Edwards, A. S., Everhart, M., Goins, Q. R., … Williams, J. D. (2017). The development of a secondary-level solo wind instrument performance rubric using the multifaceted rasch partial credit measurement model. *Journal of Research in Music Education*, 2242941769487.

Wesolowski, B. C., Wind, S. A., & Engelhard, G. (2015). Rater fairness in music performance assessment: Evaluating model-data fit and differential rater functioning. *Musicae Scientiae*, *19*(2), 147–170. http://doi.org/10.1177/1029864915589014

Wesolowski, B. C., Wind, S. A., & Engelhard, G. (2016a). Examining rater precision in music performance assessment: An analysis of rating scale structure using the multifaceted rasch partial credit model. *Music Perception*, *33*(5), 662–678.

Wesolowski, B. C., Wind, S. A., & Engelhard, G. (2016b). Rater analyses in music performance assessment : Application of the Many Facet Rasch Model. In *Connecting practice, measurement, and evaluation: Selected papers from the 5th International Symposium on Assessment in Music Education* (pp. 335–356).

Witt, A. C. (1986). Use of class time and student secondary attentiveness in instrumental music rehearsals. *Journal of Research in Music Education*, *34*(1), 34–42.

Worthy, M. D. (2006). Observations of three expert wind conductors in college rehearsals. *Bulletin of the Council for Research in Music Education*, *168*(168), 51–62.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.

Yarbrough, C. (1975). Effect of magnitude of conductor behavior on students in selected mixed choruses. *Journal of Research in Music Education*, *23*(2), 134–146.

Yarbrough, C., & Price, H. E. (1981). Prediction of performer attentiveness based on rehearsal activity and teacher behavior. *Journal of Research in Music Education*, *29*(3), 209. http://doi.org/10.2307/3344994

Yarbrough, C., & Price, H. E. (1989). Sequential patterns of instruction in music. *Journal of Research in Music Education*, *37*(3), 179–187. http://doi.org/10.2307/3344668

Younger, K. G. (1998). *An observational analysis of instructional effectiveness in intermediate level band and orchestra rehearsals*. (Doctoral dissertation, Universitiy of Texas at Austin).

Zdzinski, S. F., & Barnes, G. . (2002). Development and validation of a string performance trating scale. *Journal of Research in Music Education*, *50*, 245–255.

APPENDICES

APPENDIX A

Adjudication form from Southern California

Southern California School Band & Orchestra Association
Band and Orchestra Festival Adjudication Form

School: _____ **Location (City):** _____

Ensemble Name: _____

Director's Name: _____ **SCSBOA Number:** _____

Number of Years Teaching Experience _____ **Number of Years at this School** _____

Ensemble Classification: _____ **CIRCLE ONE:** **Playing for Rating** **Constructive Criticism Only**

Selections in Performance Order

1 **Is this selection from either the SCSBOA or Texas music list?** Y or N **If yes, which:** SCSBOA / Texas
If yes, please indicate the GRADE level of the piece: 0.5, 1, 2, 3, 4/5

Title: _____

Composer: _____

2 **Is this selection from either the SCSBOA or Texas music list?** Y or N **If yes, which:** SCSBOA / Texas
If yes, please indicate the GRADE level of the piece: 0.5, 1, 2, 3, 4/5

Title: _____

Composer: _____

3 **Is this selection from either the SCSBOA or Texas music list?** Y or N **If yes, which:** SCSBOA / Texas
If yes, please indicate the GRADE level of the piece: 0.5, 1, 2, 3, 4/5

Title: _____

Composer: _____

Note: Perform a third selection ONLY if neither selection 1 nor selection 2 is a Grade 4 or 5 selection.
This selection may NOT be a Grade 4/5 selection.

DIRECTIONS FOR DETERMINING GRADE CLASSIFICATION

Select the correct factor number in each area and add the numbers to obtain your proper grade classification.
See Classification Points Box for computing.

EXPERIENCE AND BACKGROUND
Select the correct factor numbers from the scales below

A. FESTIVAL EXPERIENCE — **FACTOR**

No festival experience, unit or feeder	1	
Festival experience, either unit or feeder	2	
Festival experience, both unit and feeder	3	_____

B. INSTRUMENTATION

Incomplete to complete	1-2-3-4-5	_____

C. MEMBERSHIP — **FACTOR**

Lower grade level (cadet or training)	0	
General (all students in the program)	1	
Upper grade level (all students in grade)	2	
Select (e.g.: by audition)	4	_____

D. PRIVATE LESSONS

None to high percentage	0-1-2-3	_____
TOTAL		_____

CLASSIFICATION SUMMARY (circle one)

Elem./Int./M.S./JHS		High School	
Grade 0.5	1 - 5	Grade 2	1 - 5
Grade 1	5 - 11	Grade 3	5 - 11
Grade 2	11 - 15	Grade 4/5	11 - 15

Grade 0.5 is intended for Elementary Ensembles and **FIRST** year Intermediate School, Middle School and Junior High School Ensembles regardless of factor total. Grades 1 and 2 are intended for Intermediate School, Middle School and Junior High School Ensembles. Selections from Grade 3 or 4 may be performed by Class 0.5-1-2 ensembles of exceptional ability.
Grades 2, 3 and 4/5 are intended for High School Ensembles, although a cadet or training ensemble may enter as a Grade 2 ensemble regardless of factor total.

APPENDIX B

Adjudication form from Georgia

BAND LARGE GROUP PERFORMANCE EVALUATION

Date: _____ **Classification:** _____ **No of Players:** _____

School: _____

Name of Performing Group: _____

Selections: **1.** _____

2. _____

3. _____

Final Rating

Use of no Plus or Minus in final rating

Adjudicator will grade principal items **A, B, C, D or E** or numerals in the respective squares **for each selection**. Comments must deal with fundamental principals and be constructive. Minor details may be marked on music furnished to adjudicators.

	1.	2.	3.
TONE (beauty, blend, control)			
INTONATION (chords, melodic line, tutti)			
TECHNIQUE (articulation, facility, precision, rhythm)			
BALANCE (ensemble, sectional)			
MUSICALITY (expression, phrasing, style, tempo, artistry, fluency)			
OVERALL			

*May be continued on other side

OTHER FACTORS (not included in the graded evaluation)
Choice of music, appearance, stage presence

Signature of Adjudicator: _____

APPENDIX C

Adjudication form from Iowa

IOWA HIGH SCHOOL MUSIC ASSOCIATION
LARGE GROUP ADJUDICATION FORM

Performance Time _____ Date _____

Event _____ Festival Site _____ Class _____
Bass Clef Chorus, Concert Band, etc.

School _____ Director Name _____ No. in Ensemble _____

Selections	Composer	Publisher
1.		
2.		
3.		

BANDS AND CHORUSES, PLEASE (*) YOUR REQUIRED SELECTION

Place one of these numbers in each circle below, then total carefully.

5 - A superior performance-outstanding in nearly every detail
4 - An excellent performance-minor defects.
3 - A good performance-lacking finesse and/or interpretation.
2 - A fair performance-basic weaknesses.
1 - A poor performance -unsatisfactory

AREAS OF CONCERN	COMMENTS
◯ **Tone Quality** Consider: resonance, control, clarity, focus, consistency, warmth	
◯ **Intonation** Consider: within ensemble, accuracy to printed pitches	
◯ **Rhythm** Consider: accuracy of note and rest values, duration, pulse, steadiness, correctness of meters	
◯ **Balance, Blend** Consider: Likeness of qualities, awareness of ensemble, accompaniment	
◯ **Technique** (facility/accuracy) Consider: artistry, attacks, releases control of ranges, musical and/or mechanical skill	
◯ **Interpretation, Musicianship** Consider: style, phrasing, tempo, dynamics, emotional involvement	
◯ **Diction - Vocal** **Bowing - Strings** **Articulation - Winds**	
◯ **Other Performance Factors** Consider: choice of literature, appropriate appearance, poise, posture, general conduct, mannerisms, facial expression (vocal), memory (if required)	

◯ **TOTAL POINTS**

_____ **DIVISIONAL RATING**

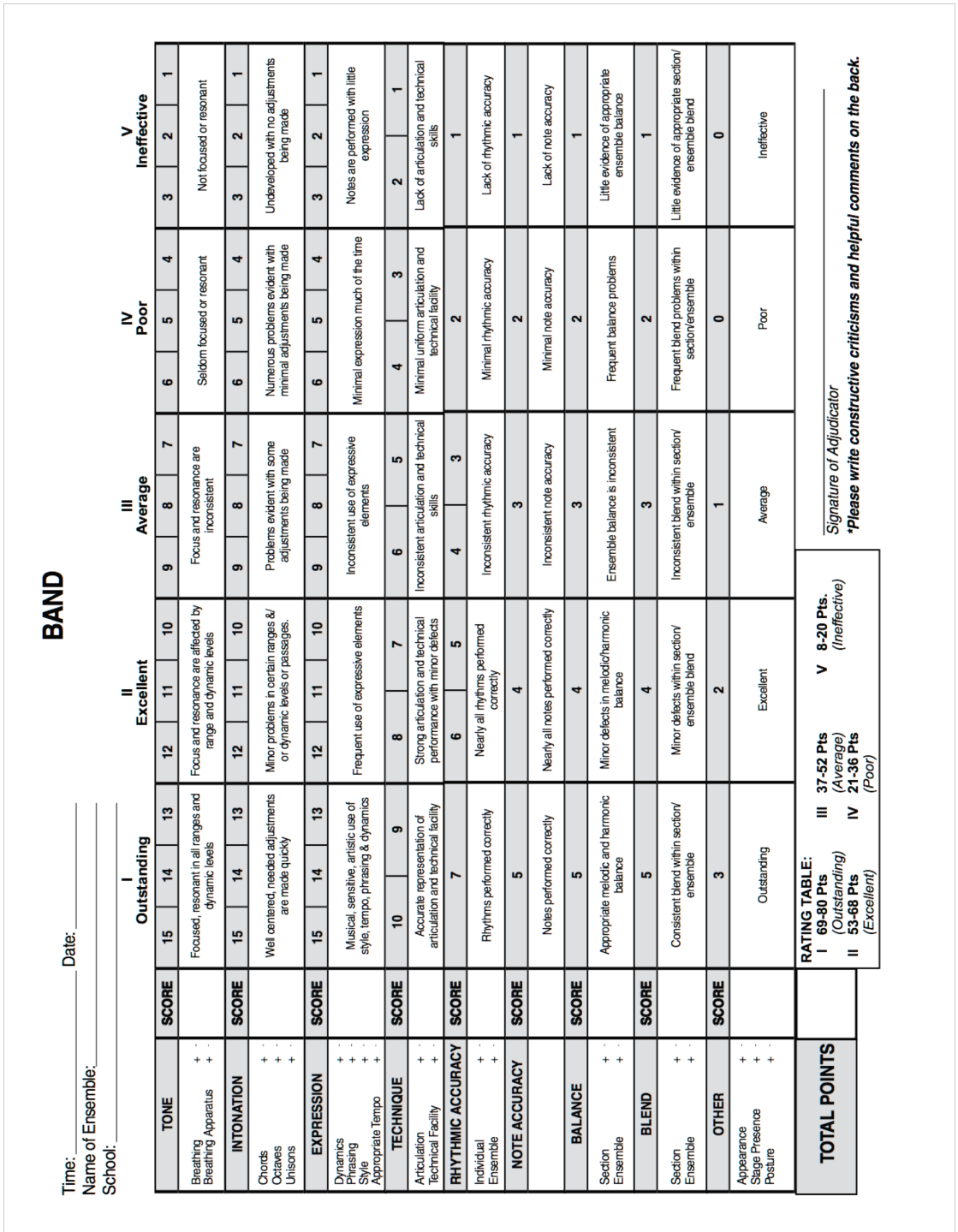
RATING COMPUTATION TABLE

40-36 points = Division I (Superior)
35-30 points = Division II (Excellent)
29-22 points = Division III (Good)
21-12 points = Division IV (Fair)
11- 8 points = Division V (Poor)

_____
Adjudicator Signature

APPENDIX D

Adjudication form from Kansas



BAND

Time: _____ Date: _____
Name of Ensemble: _____
School: _____

		I Outstanding			II Excellent			III Average			IV Poor			V Ineffective		
TONE	**SCORE**	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Breathing Breathing Apparatus	+ - + -	Focused, resonant in all ranges and dynamic levels			Focus and resonance are affected by range and dynamic levels			Focus and resonance are inconsistent			Seldom focused or resonant			Not focused or resonant		
INTONATION	**SCORE**	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Chords Octaves Unisons	+ - + - + -	Well centered, needed adjustments are made quickly			Minor problems in certain ranges &/ or dynamic levels or passages.			Problems evident with some adjustments being made			Numerous problems evident with minimal adjustments being made			Undeveloped with no adjustments being made		
EXPRESSION	**SCORE**	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Dynamics Phrasing Style Appropriate Tempo	+ - + - + - + -	Musical, sensitive, artistic use of style, tempo, phrasing & dynamics			Frequent use of expressive elements			Inconsistent use of expressive elements			Minimal expression much of the time			Notes are performed with little expression		
TECHNIQUE	**SCORE**	10		9	8		7	6		5	4		3	2		1
Articulation Technical Facility	+ - + -	Accurate representation of articulation and technical facility			Strong articulation and technical performance with minor defects			Inconsistent articulation and technical skills			Minimal uniform articulation and technical facility			Lack of articulation and technical skills		
RHYTHMIC ACCURACY	**SCORE**	7			6		5	4		3	2			1		
Individual Ensemble	+ - + -	Rhythms performed correctly			Nearly all rhythms performed correctly			Inconsistent rhythmic accuracy			Minimal rhythmic accuracy			Lack of rhythmic accuracy		
NOTE ACCURACY	**SCORE**	5			4			3			2			1		
		Notes performed correctly			Nearly all notes performed correctly			Inconsistent note accuracy			Minimal note accuracy			Lack of note accuracy		
BALANCE	**SCORE**	5			4			3			2			1		
Section Ensemble	+ - + -	Appropriate melodic and harmonic balance			Minor defects in melodic/harmonic balance			Ensemble balance is inconsistent			Frequent balance problems			Little evidence of appropriate ensemble balance		
BLEND	**SCORE**	5			4			3			2			1		
Section Ensemble	+ - + -	Consistent blend within section/ ensemble			Minor defects within section/ ensemble blend			Inconsistent blend within section/ ensemble			Frequent blend problems within section/ensemble			Little evidence of appropriate section/ ensemble blend		
OTHER	**SCORE**	3			2			1			0			0		
Appearance Stage Presence Posture	+ - + - + -	Outstanding			Excellent			Average			Poor			Ineffective		
TOTAL POINTS																

RATING TABLE:
I 69-80 Pts *(Outstanding)*
II 53-68 Pts *(Excellent)*
III 37-52 Pts *(Average)*
IV 21-36 Pts *(Poor)*
V 8-20 Pts. *(Ineffective)*

Signature of Adjudicator

Please write constructive criticisms and helpful comments on the back.

APPENDIX E

Adjudication form from Nebraska

JUDGE'S COMMENTS

TONE
(1-15)

INTONATION
(1-15)
EXPRESSION

(1-15)
TECHNIQUE
(1-10)

RHYTHMIC
ACCURACY
(1-7)

NOTE
ACCURACY
(1-5)
BALANCE

(1-5)
BLEND
(1-5)
OTHER

(0-3)
GENERAL

BAND / WIND INSTRUMENT ENSEMBLE

(Large ensembles require three forms per entry.)

Order or Time of Appearance:_______ Event #:______ Class:_____ Date:________

School:____

Name of Group:______________________________________

Selections:____

____

Rating
7-10, Div I **11-17, Div II** **18-24, Div III** **25-31, Div IV** **32+, Div V**

Score captions below 1 to 5 with 1 being <u>highest</u> score earned. Convert total points scored to scale at right for final rating.

☐ ***TONE*** *(beauty, blend, control)* **1 - Maturely consistent in concept and control 2 - Mostly strong with occasional inconsistencies, 3 - Repeated, distracting inconsistencies 4 - Lacking clear tonal concepts 5 - Fundamentally missed**

☐ ***INTONATION*** (*tonality awareness, pitch center, unisons/intervals/chords)* **1 - Consistent awareness/concept of key center 2 - Strong; Occasional inconsistencies 3 - Inaccurate unisons and harmonies 4 - Consistently missing key signatures 5 - Lacking fundamental awareness**

☐ ***PRECISION*** *(accuracy to the printed page - rhythms/notes/attacks/releases/unified breathing)* **1 - Consistent, detailed accuracy regarding rhythms, notes, attacks, releases, etc, 2 - Strong with occasional inconsistencies 3 - Distractingly inconsistent, 4 - Unclear with technical concepts 5 - Fundamentally unsound**

☐ ***FACILITY*** (*ease of performing bowings / fingerings / positions / shifts; articulations, phrasing, dexterity)* **1 - Confident throughout, 2 - Strong with occasional inconsistencies, 3 - Tentative presentation throughout, 4 - Insecurities repeatedly distract performance, 5 - Fundamentally unsound**

☐ ***BALANCE*** (*ensemble / sectional; melodies / harmonies*) **1 - All voicings consistently and effectively presented 2 - Strong with occasional inconsistencies 3 - Repeated missed voicings 4 - Unclear concept of balance 5 - Concept of balance missing**

☐ ***EXPRESSION / INTERPRETATION*** (*phrase shape/musical line, contrasts in tempos/dynamics/articulations*) **1 - Consistently and effectively applying printed and implied musical design 2 - Strong, but inconsistent 3 - Little variety in presentation 4 - Minimal variety presented 5 - None present**

☐ ***OTHER FACTORS*** (*Choice of Music, ensemble feel, discipline, posture, stage presence, appearance, instrumentation*) **1 - Outstanding literature and presentation 2 - Strong with some questions 3 - Multiple distracting elements 4 - Little or no attention to details 5 - Unacceptable presentation**

_______ **TOTAL POINTS** **Signature of Adjudicator**: ______________________________

[Final Rating Conversion Scale below Rating Box at top right of ballot]

Revised 5/2011

APPENDIX F

Adjudication form from North Carolina

North Carolina Bandmasters Association

Music Performance Adjudication

Concert Band - Stage Form

Final Rating ☐

Name of Ensemble _____ **School** _____

Date	Time	No. of Members	District	Performance Grade

Adjudication Scale By Rating: I = Superior II = Excellent III = Average IV = Below Average V = Poor

Repertoire: Title (checkbox to left to indicate March) — **Composer / Arranger** — **Grade**

☐ _____ _____ _____

☐ _____ _____ _____

☐ _____ _____ _____

Grade each caption using A, B, C, D or F in its respective box
Comments should be constructive. It is acceptable to use + / – in the individual captions.

Tone Quality: Characteristic Sound, Clarity, Consistency, Control, Matching within Section, Resonance ☐

_____

_____

_____

Intonation: Initial pitch, Chords, Individual, Melodic line, Makes Adjustments ☐

_____

_____

_____

Balance: Blend, Ensemble, Melodic, Section ☐

_____

_____

_____

Precision: Accuracy, Articulation, Entrances, Releases, Pulse, Rhythm, Technique ☐

_____

_____

_____

Musicianship: Adherence to printed musical directions not addressed by previous captions ☐

_____

_____

_____

Interpretation: Choices of Tempos, Dynamic Contrast, Phrasing, Style, Energy, Expression. ☐

_____

_____

_____

General Factors: Choice of Appropriate Literature, Instrumentation, Etiquette, Confidence, Discipline, Appearance, Posture ☐

_____

_____

_____

Signature of Adjudicator: _____

APPENDIX G

Adjudication form from Texas

UNIVERSITY INTERSCHOLASTIC LEAGUE

4

CONCERT

Entry Blank and Comment Sheet

Please read the current issue of the Constitution and Contest Rules

School_____ **City**_____ **Conference**_____ **Organization Event**_____

Contest Date_____ **Performing Group Type**_____ __ __ **Number of Students**____ **Region**___

Director _____ **Additional Directors**_____

Composer/Arranger	Title of selections and movements	UIL ID#
_____	_____	_____
_____	_____	_____
_____	_____	_____

TONE
+ - Centered, focused tone quality
+ - Balance within sections
+ - Balance between sections
+ - Intonation within sections
+ - Intonation between sections
+ - Dynamic contrasts without distortion

TECHNIQUE
+ - Note Accuracy
+ - Manual dexterity and flexibility
+ - Rhythmic accuracy
+ - Rhythmic stability
+ - Appropriate mastery of articulation
+ - Observance of ties, slurs and articulation markings

MUSICIANSHIP
+ - Appropriateness of style
+ - Sensitivity to phrasing
+ - Observance of musical markings
+ - Appropriateness of dynamic contrasts
+ - Appropriate observance of tempo
+ - Demonstrates musical understanding

OTHER COMMENTS (No rating applies)

FINAL RATING **I** **II** **III** **IV** **V**
Write in rating here _____ Signature of official_____

APPENDIX H

Adjudication form from Virginia

Instrumental Ensemble - Wind

Rating

Order or time of appearance _____ Event No. _____ Class _____ Date _____

Name of Ensemble _____

Performer(s) Name(s) _____

School _____ City _____ State _____ District _____

Selection _____ Instrument _____

Adjudicator will grade principal items A, B, C, D or F in the respective squares. Comments Must deal with fundamental principles and be constructive. Minor details may be marked on music.

TONE (beauty, blend, control) _____

INTONATION (harmonic parts, melodic line, tutti) _____

TECHNIQUE (articulation, embouchure, facility, precision, rhythm) _____

BALANCE _____

INTERPRETATION (expression, phrasing, style, tempo) _____

MUSICAL EFFECT (artistry, fluency) _____

OTHER FACTORS (choice of music, stage presence and appearance) _____

Signature of Adjudicator _____

BAND PERFORMANCE RATING SCALE

Instructions: Please respond to each of the following statements on the basis of how strongly you agree or disagree that the statement is descriptive of the performance. Use the following five-point scale:

SA - Strongly agree that the statement describes the performance
A - Slightly agree that the statement describes the performance
NA - Neither agree nor disagree that the statement describes the performance
D - Slightly disagree that the statement describes the performance
SD - Strongly disagree that the statement describes the performance

Please choose only one response to each statement, but be sure to answer each item. Then circle your choice.

SA A NA D SD 1. Intonation is quite good.

SA A NA D SD 2. Basic tuning is not good, band plays out of tune.

SA A NA D SD 3. Tone is shallow.

SA A NA D SD 4. The band plays with a well defined pitch center.

SA A NA D SD 5. The band has good control in high pitch registers.

SA A NA D SD 6. Tutti chords are badly out of tune.

SA A NA D SD 7. Excellent balance between parts in a section.

SA A NA D SD 8. Excellent balance in all parts.

SA A NA D SD 9. Lower parts balance the group well.

SA A NA D SD 10. All sections are well balanced within.

SA A NA D SD 11. Inner parts balance well.

SA A NA D SD 12. Inner parts are too timid.

SA A NA D SD 13. Performance shows careful attention to dynamics.

SA A NA D SD 14. Performance has a good, wide variety in dynamics.

SA A NA D SD 15. Needs to be more expressive.

SA A NA D SD 16. Crescendo and diminuendo are properly graduated.

SA A NA D SD 17. Performance lacks emotion.

SA A NA D SD 18. Performance exhibits proper style.

SA A NA D SD 19. Dotted rhythms played as triplets.

SA A NA D SD 20. Dotted eighth-sixteenth pattern is inaccurate.

SA A NA D SD 21. Syncopated patterns are correctly played.

SA A NA D SD 22. Rhythmically accurate.

SA A NA D SD 23. Performance is marked by unsteady rhythm.

SA A NA D SD 24. Rhythmic figures are properly and distinctly executed.

SA A NA D SD 25. Runs are played accurately and smoothly.

SA A NA D SD 26. Notes in runs are inaccurate.

SA A NA D SD 27. Fingering problems cause a lack in clarity.

SA A NA D SD 28. Awkward and difficult passages are not prepared.

SA A NA D SD 29. Entrances are not precise.

SA A NA D SD 30. Technique is inadequate for performance of this music.

APPENDIX J

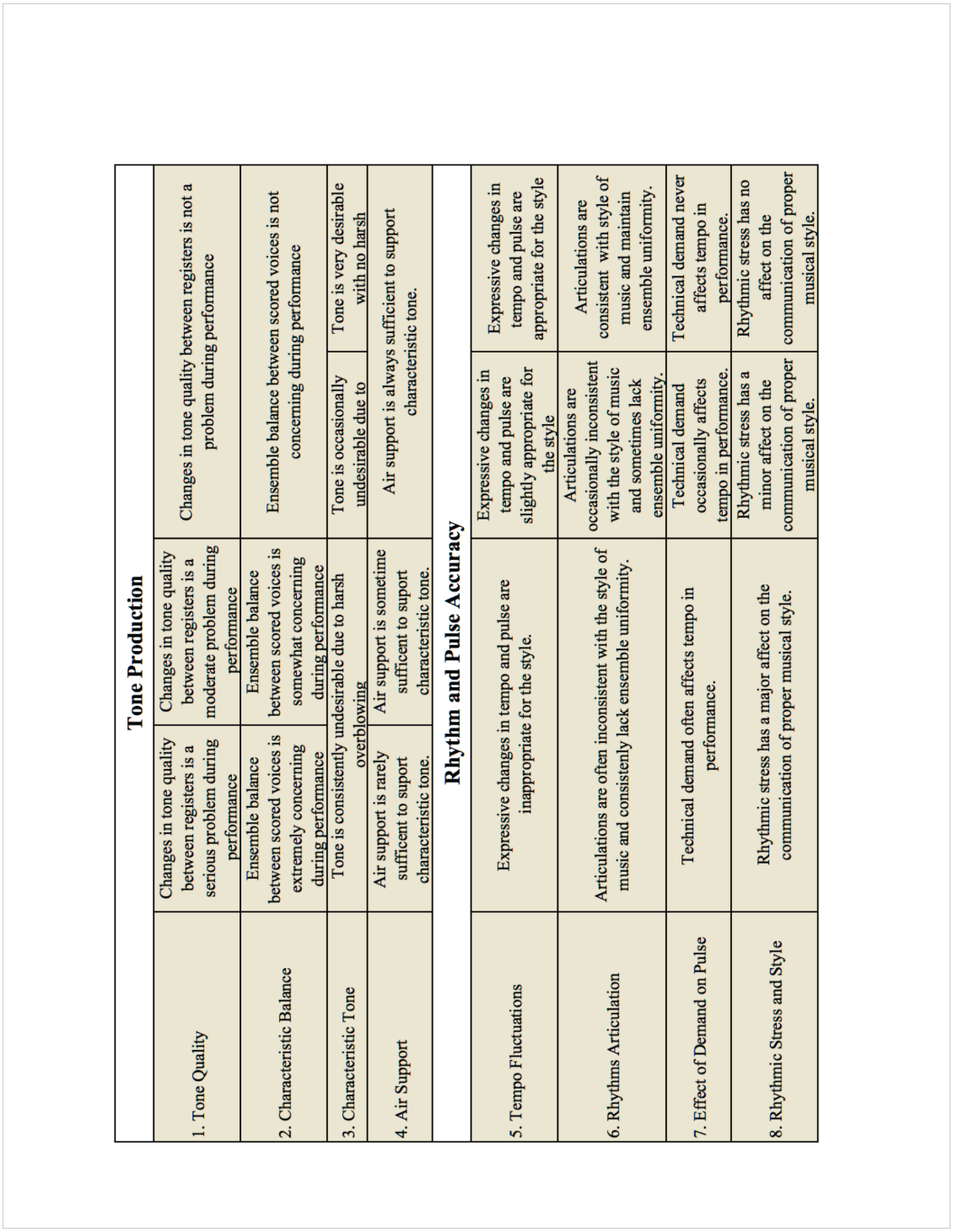
Ensemble Fundamentals Survey

A. How many days per week do you typically rehearse					
Possible Responses	1	2	3	4	5
B. How frequently do you engage in the following types of ensemble fundamentals?					
Fundamental Type	Possible Responses				
Long Tones	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Lip Slurs	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Scales/Arpeggios	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Tuning Exercises	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Chorales	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Sight-Reading	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Rhythm Drills	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Articulation Exercises	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Breathing Exercises	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Dynamics Exercises	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
Balance/Blend Exercises	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never

Singing	Every Rehearsal	More than half of all rehearsals	Less than half of all rehearsals	Once per week or less	Never
C. How much time on average do you spend on each type of ensemble fundamental when you use it?					
Fundamental Type	Possible Responses				
Long Tones	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Lip Slurs	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Scales/Arpeggios	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Tuning Exercises	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Chorales	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Sight-Reading	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Rhythm Drills	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Articulation Exercises	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Breathing Exercises	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Dynamics Exercises	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Balance/Blend Exercises	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
Singing	Less than 1 minute	1-3 minutes	3-5 minutes	5 minutes or more	
D. How much total time do you spend on fundamentals in a typical rehearsal?					
Possible Responses	None	5 minutes or less	5-10 minutes	10-15 minutes	15 minutes or more

APPENDIX K

Edwards Band Rating Rubric



Tone Production				
1. Tone Quality	Changes in tone quality between registers is a serious problem during performance	Changes in tone quality between registers is a moderate problem during performance	Changes in tone quality between registers is not a problem during performance	
2. Characteristic Balance	Ensemble balance between scored voices is extremely concerning during performance	Ensemble balance between scored voices is somewhat concerning during performance	Ensemble balance between scored voices is not concerning during performance	
3. Characteristic Tone	Tone is consistently undesirable due to overblowing		Tone is occasionally undesirable due to	Tone is very desirable with no harsh
4. Air Support	Air support is rarely sufficient to suport characteristic tone.	Air support is sometime sufficent to suport characteristic tone.	Air support is always sufficient to support characteristic tone.	
Rhythm and Pulse Accuracy				
5. Tempo Fluctuations	Expressive changes in tempo and pulse are inappropriate for the style.		Expressive changes in tempo and pulse are slightly appropriate for the style	Expressive changes in tempo and pulse are appropriate for the style
6. Rhythms Articulation	Articulations are often inconsistent with the style of music and consistenly lack ensemble uniformity.		Articulations are occasionally inconsistent with the style of music and sometimes lack ensemble uniformity.	Articulations are consistent with style of music and maintain ensemble uniformity.
7. Effect of Demand on Pulse	Technical demand often affects tempo in performance.		Technical demand occasionally affects tempo in performance.	Technical demand never affects tempo in performance.
8. Rhythmic Stress and Style	Rhythmic stress has a major affect on the communication of proper musical style.		Rhythmic stress has a minor affect on the communication of proper musical style.	Rhythmic stress has no affect on the communication of proper musical style.

9. Steady Pulse	Control of pulse detracts much from the continuous flow of the music		Contol of pulse sometimes detracts from the continuous flow of the music.	Control of pulse does not detract from the continuous flow of the music.
10. Ensemble Releases	Ensemble releases are almost never executed with precision across performers.		Ensemble releases are sometimes executed with precision across performers.	Ensemble releases are almost always executed with precision across performers.
11. Rhythmic Subdivision	Inaccurate performance of subdivisions frequently detract from solidly communicated tempo and meter.		Inaccurate performance of subdivisions occasionally detract from solidly communicated tempo and meter.	Accurate performance of subdivisions contribute to solidly communicated tempo and meter.
Pitch and Intonation Accuracy				
12. Intonation in Expression	Dynamic fluctuations often detract from proper intonation.	Dynamic fluctuations sometimes detract from proper intonation.	Dynamic fluctuations rarely detract from proper intonation.	
13. Key Signatures	Inaccurate performance of key signatures often detract from performance.		Inaccurate performance of key signatures occasionally detract from performance.	Accurate performance of key signature support performance.
14. Technical Efficiency	Technical efficiency is an extreme barrier to the quality of the performance.	Technical efficiency is a moderate barrier to the quality of the performance.	Technical efficiency somewhat of a barrier to the quality of the performance.	Technical efficiency is not a barrier to the quality of the performance.
15. Intonation	Intonation accuracy is a serious problem.	Intonation accuracy is a moderate problem.	Intonation accuracy is not a problem.	
Expressive Qualities / Stylistic Interpretations				
16. Dynamics	Ensemble rarely demonstrates meaningful contrast in dynamics.		Ensemble sometimes demonstrates meaningful contrast in dynamics.	Ensemble frequently demonstrates meaningful contrast in dynamics.

17. Melody verses Accompaniment	Proper balance of the ensemble is an extreme problem in performance.	Proper balance of the ensemble is a moderate problem in performance.	Proper balance of the ensemble is a minor problem in performance.	Proper blanance of the ensemble not a problem in performance.
18. Cadential Inflection	Phrases lack proper inflection at cadential points.		Phrases demonstrate proper inflection at cadential points.	
19. Articulations Style	Articulations are often inconsistent in passages with notes of a similar style and detract much from the performance.	Articulations are occasionally inconsistent in passages with notes of a similar style and slightly detract from the performance.	Articulations are consistent in passages with notes of a similar style and do not detract from the performance.	
20. Expressive Gestures	Stylistic or expressive modifications are rarely appropriate or present in performance		Stylistic or expressive modifications are typically appropriate and somewhat present in performance.	Sytylistic or expressive modifications are appropriate and consistently present in performance.
21. Musical Style	Appropriate musical style is rarely communicated through the ensembles performance		Appropriate musical style is sometimes communicated through the ensembles performance	Appropriate musical style is consistencly communicated through the ensembles performance.
22. Ensemble Balance	Individual ensemble members create an extreme barrier to proper ensemble balance.	Individual ensemble members create somewhat of a barrier to proper ensemble balance.	Individual enesmble members do not create a barrier to proper ensemble blanance.	
23. Effect of Demand on Balance	Technical demand is an extreme barrier to proper ensemble balance.		Technical demand is somewhat of a barrier to proper ensemble balance.	Technical demand is not a barrier to proper ensemble balance.

APPENDIX L

IRB Approval

Phone 706-542-3199

The University of Georgia

Office of the Vice President for Research
Institutional Review Board

APPROVAL OF PROTOCOL

March 2, 2016

Dear Brian Wesolowski:

On 3/2/2016, the IRB reviewed the following submission:

Type of Review:	Initial Study
Title of Study:	The Relationship of Ensemble Fundamentals and Perceived Performance Quality of High School Bands
Investigator:	Brian Wesolowski
IRB ID:	STUDY00003074
Funding:	None
Grant ID:	None

The IRB approved the protocol from 3/2/2016.

In conducting this study, you are required to follow the requirements listed in the Investigator Manual (HRP-103).

Sincerely,

Dr. Gerald E. Crites, MD, MEd
University of Georgia
Institutional Review Board Chairperson

APPENDIX M

Calibration of Raters

Rater Number	Observed Average	Measure	SE	Infit *MSE*	Std. Infit	Outfit *MSE*	Std. Outfit
17	3.00	3.89	0.40	1.14	0.50	3.39	2.00
13	2.40	2.65	0.24	1.13	0.70	0.96	0.00
16	2.50	2.32	0.24	0.96	-0.20	0.90	-0.30
18	2.80	2.09	0.26	0.64	-2.40	0.49	-1.70
11	2.70	2.03	0.29	1.42	1.70	1.95	1.90
29	2.60	1.65	0.22	0.79	-1.60	0.74	-1.20
3	2.70	1.42	0.24	0.87	-0.90	0.73	-1.1
34	2.50	0.88	0.22	0.77	-1.60	0.82	-0.90
14	2.0	0.68	0.22	1.14	1.00	1.12	0.70
12	2.50	0.67	0.23	1.16	1.10	0.97	0.00
19	2.10	0.61	0.21	0.90	-0.60	0.77	-1.30
1	2.60	0.47	0.22	1.00	0.00	1.20	0.90
25	2.50	0.34	0.21	1.01	0.00	1.02	0.10
27	2.30	0.24	0.21	1.39	2.30	1.50	2.50
10	2.70	0.20	0.23	0.97	-0.10	0.93	-0.20
15	2.0	0.02	0.1	1.00	0.00	0.97	0.00
32	2.30	-0.10	0.21	0.83	-1.20	0.80	-1.30
7	2.60	-0.31	0.22	0.74	-2.00	0.70	-1.6
21	1.90	-0.36	0.20	1.05	0.40	1.03	0.20
24	2.20	-0.44	0.20	1.10	0.70	1.09	0.60
35	2.30	-0.55	0.21	0.93	-0.40	1.13	0.80
23	2.20	-0.56	0.20	1.54	3.10	1.48	2.80
31	2.14	-0.60	0.20	1.10	0.70	1.07	0.50
22	1.80	-0.70	0.21	0.66	-2.70	0.65	-2.50
36	2.20	-0.76	0.21	0.69	-2.50	0.68	-2.40
26	2.20	-0.85	0.20	0.76	-1.70	0.70	-2.10
28	2.00	-0.89	0.20	0.64	-2.90	0.66	-2.70
20	1.70	-0.94	0.21	0.92	-0.60	0.91	-0.50
5	2.30	-1.01	0.21	1.74	4.20	1.76	3.90
8	2.40	-1.26	0.21	0.79	-1.60	-0.74	-0.180
9	2.30	-1.43	0.20	1.29	1.90	1.42	2.50
6	2.10	-1.55	0.20	1.15	1.00	1.18	1.1
4	2.0	-1.65	0.20	0.57	-3.40	0.57	-3.30
30	1.80	-1.69	0.20	0.85	-1.00	0.86	-1.00
2	2.00	-2.02	0.20	0.83	-1.20	-0.81	-1.30
33	1.80	-2.47	0.21	1.40	2.40	1.34	1.80
Mean	2.30	0.00	0.22	1.00	-0.20	1.06	-0.20
SD	0.30	1.40	0.03	0.27	1.80	0.51	1.70

Note. The items are presented in Measure order, from most severe to least severe.

APPENDIX N

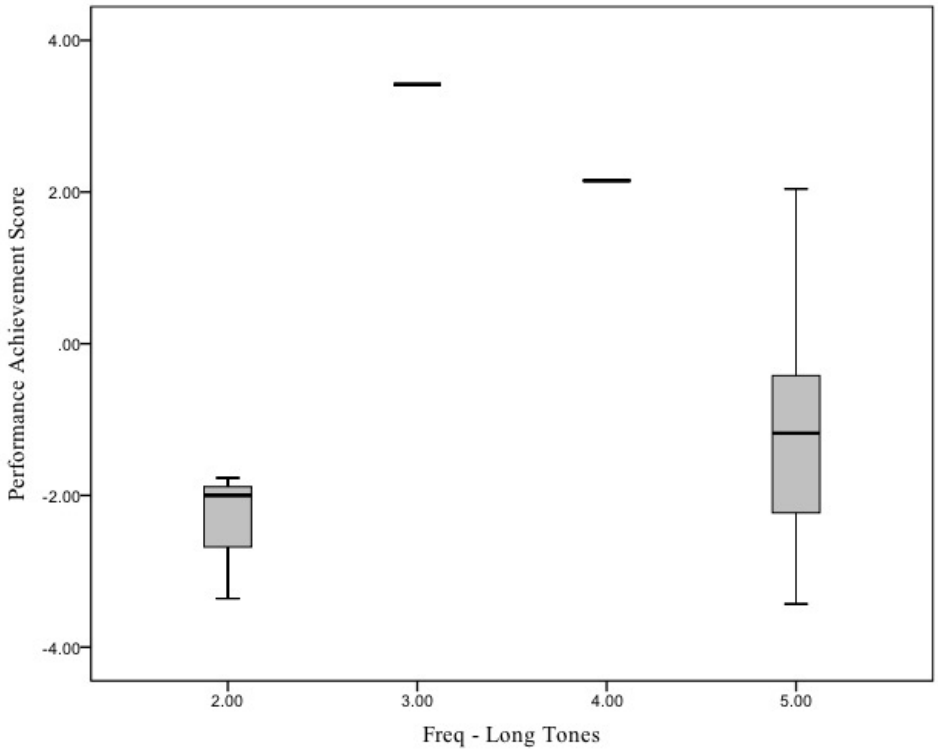
Calibration of Items

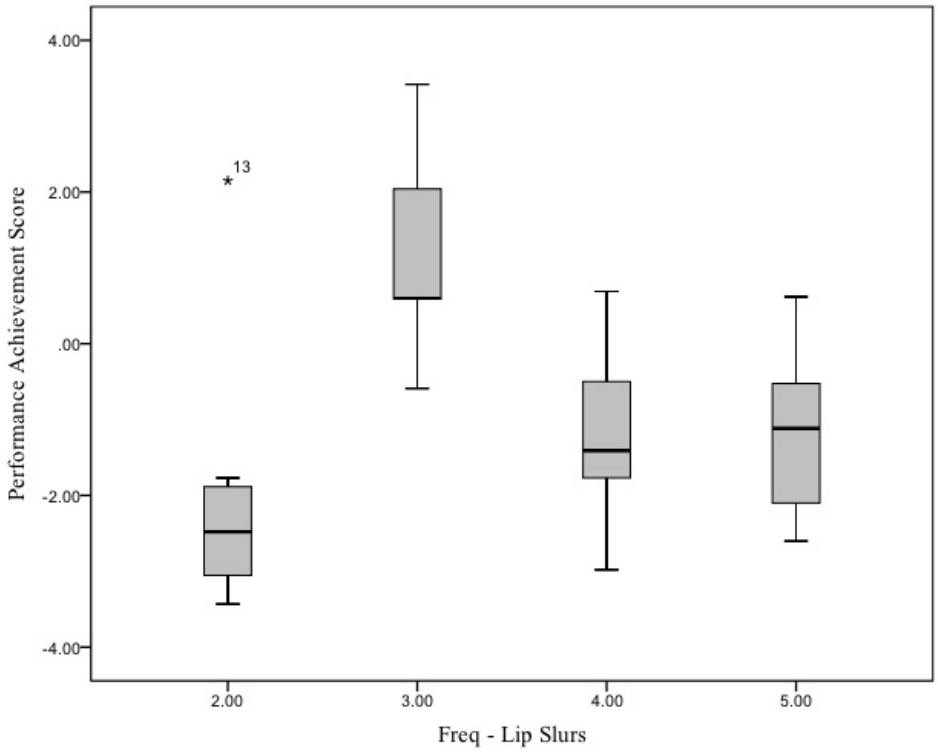
Items Number	Observed Average	Measure	SE	Infit *MSE*	Std. Infit	Outfit *MSE*	Std. Outfit
15	1.80	1.71	0.17	0.83	-1.50	0.81	-1.50
12	1.90	1.33	0.16	1.04	0.30	1.02	0.20
19	2.10	0.54	0.16	0.95	-0.40	0.88	-0.80
8	2.20	0.44	0.17	1.49	3.80	1.77	4.60
1	2.20	0.40	0.17	1.03	0.30	1.08	0.60
4	2.20	0.40	0.17	1.02	0.10	0.96	-0.20
6	2.20	0.36	0.17	0.98	-0.10	0.95	-0.30
3	2.10	0.32	0.19	0.96	-0.20	0.93	-0.40
10	2.20	0.30	0.18	1.07	0.60	1.10	0.70
18	1.60	0.09	0.22	0.88	-1.10	0.92	-0.10
17	2.90	0.07	0.15	0.90	-0.80	0.87	-1.00
14	2.80	0.03	0.14	0.95	-0.30	0.91	-0.60
16	2.30	0.00	-0.17	1.29	2.40	1.34	1.90
22	2.30	-0.02	0.17	1.01	0.10	0.99	0.00
11	2.30	-0.04	0.17	0.85	-1.30	0.78	-1.40
20	2.30	-0.08	0.18	0.92	-0.60	0.92	-0.40
23	2.30	-0.21	0.17	0.95	-0.30	0.92	-0.40
2	2.30	-0.25	0.18	0.89	-0.90	0.84	-1.00
7	2.40	-0.43	0.18	0.91	-0.70	0.88	-0.60
9	2.40	-0.70	0.18	0.88	-1.00	2.52	5.00
21	2.40	-0.84	0.18	1.00	0.00	0.95	-0.10
5	2.60	-1.69	0.20	0.93	-0.50	1.13	0.50
13	2.70	-1.74	0.20	0.98	0.00	0.82	-0.30
Mean	2.20	0.22	0.18	0.98	-0.20	0.95	-0.30
SD	0.10	0.27	0.01	0.06	0.50	0.09	0.60

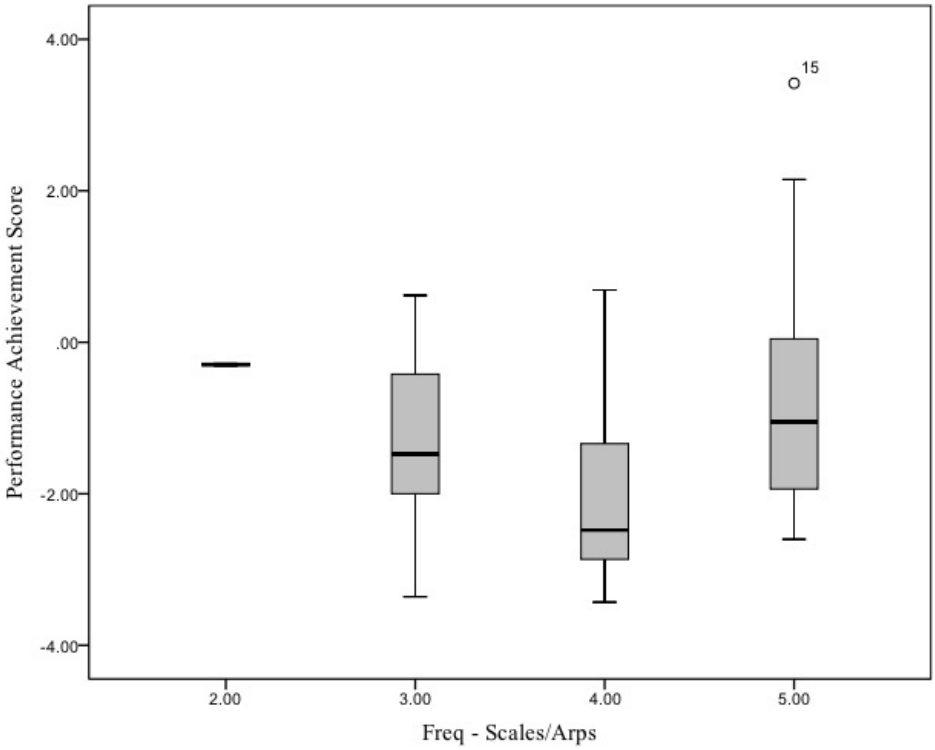
Note. The items are presented in Measure order, from most difficult to least difficult.

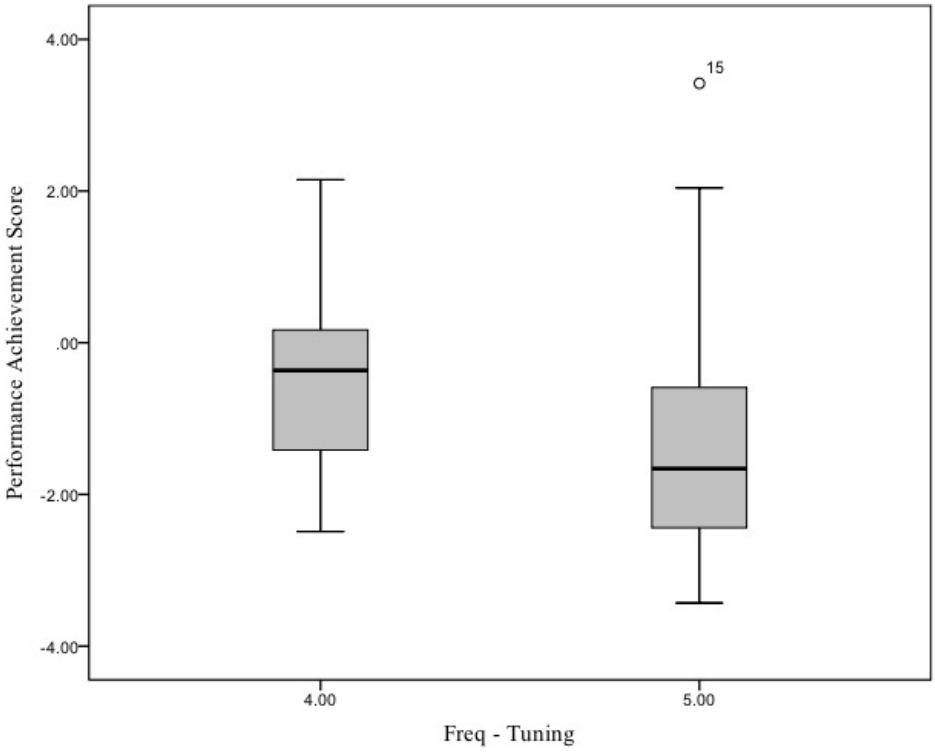
APPENDIX O

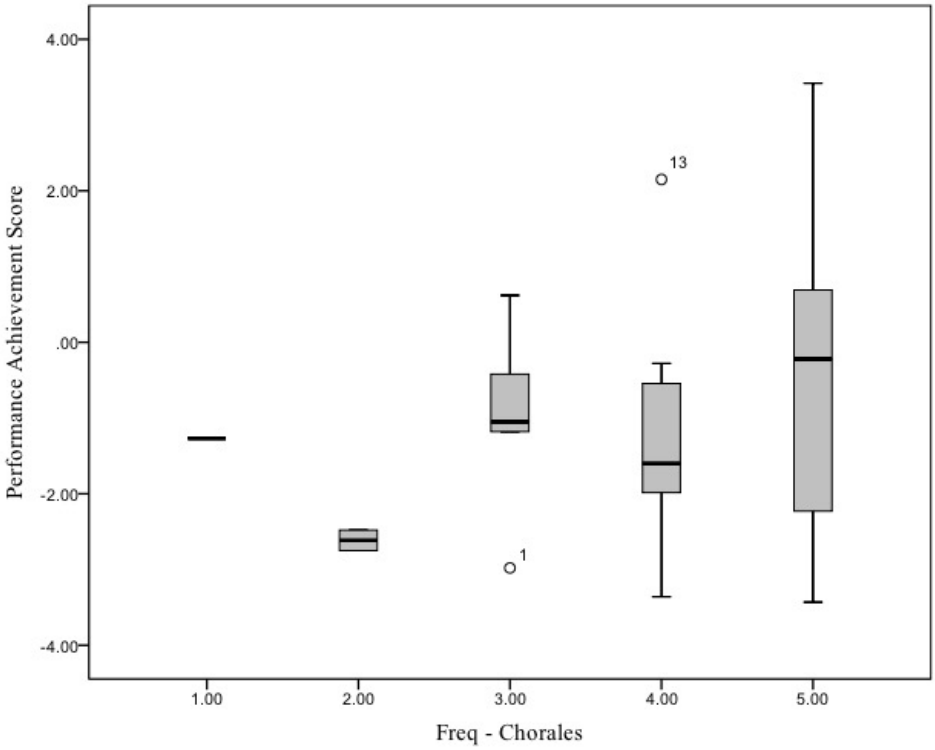
Box Plots Pertaining to Research Question 1

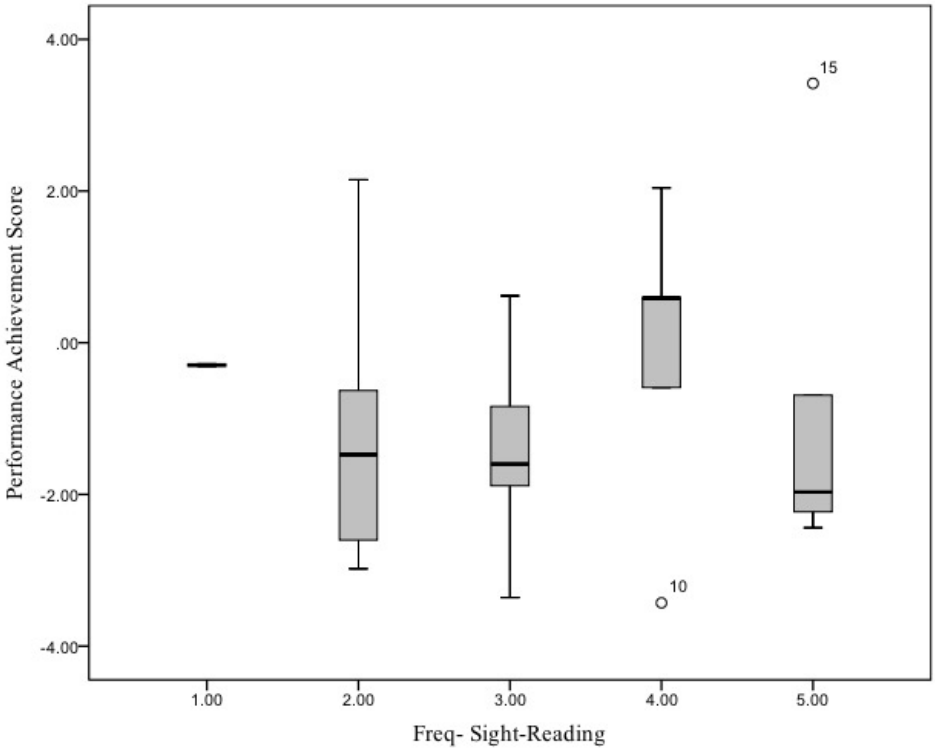


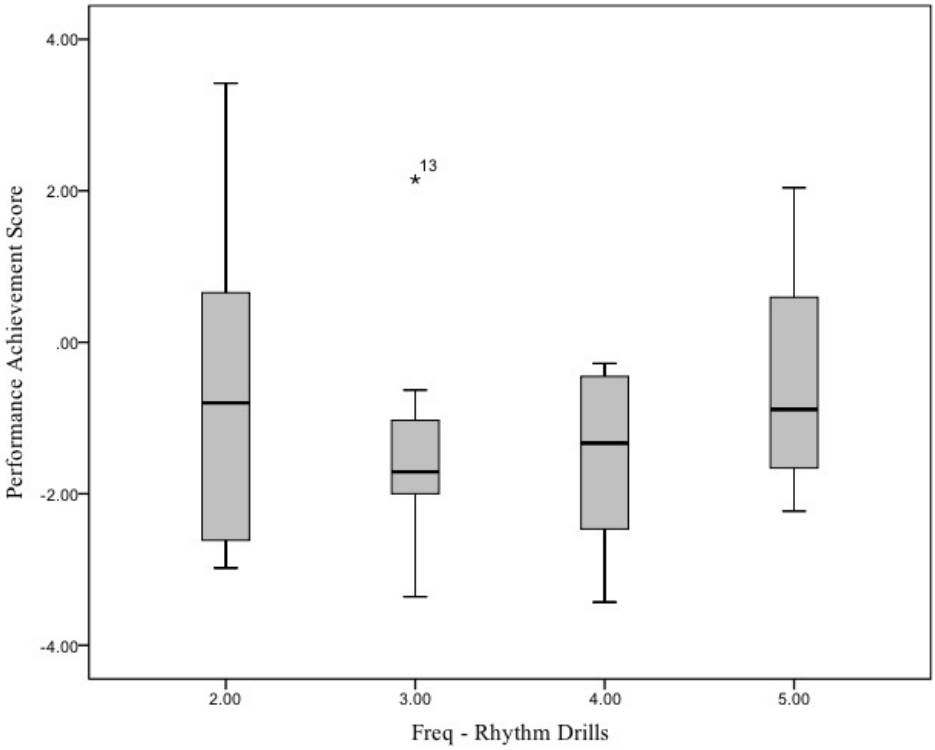


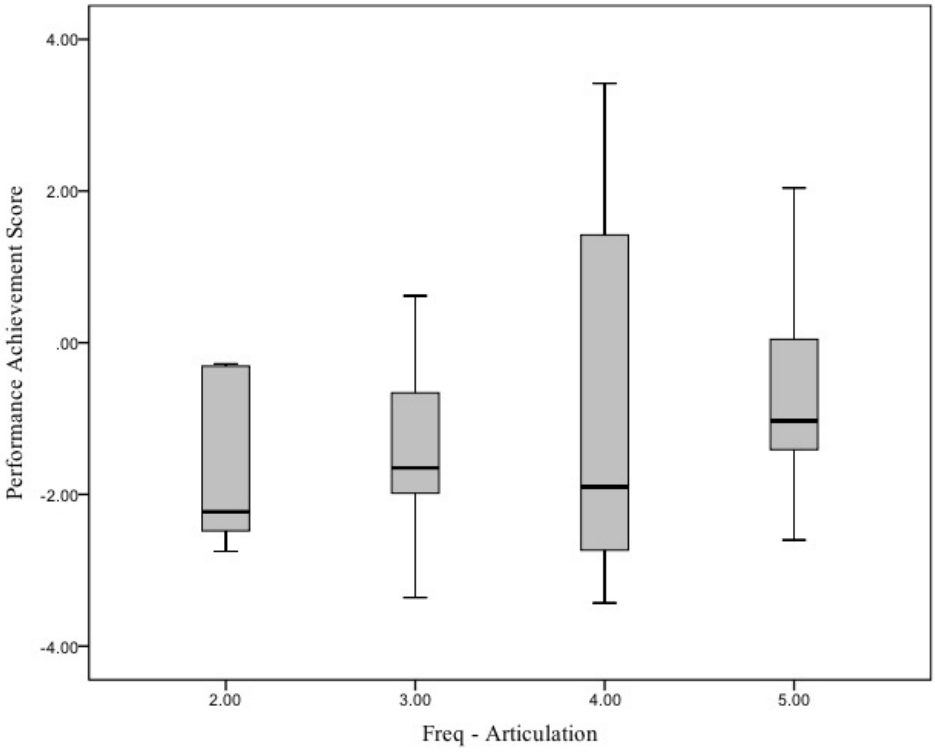

Performance Achievement Score
4.00
2.00
.00
-2.00
-4.00
15
2.00
3.00
4.00
5.00
Freq - Scales/Arps

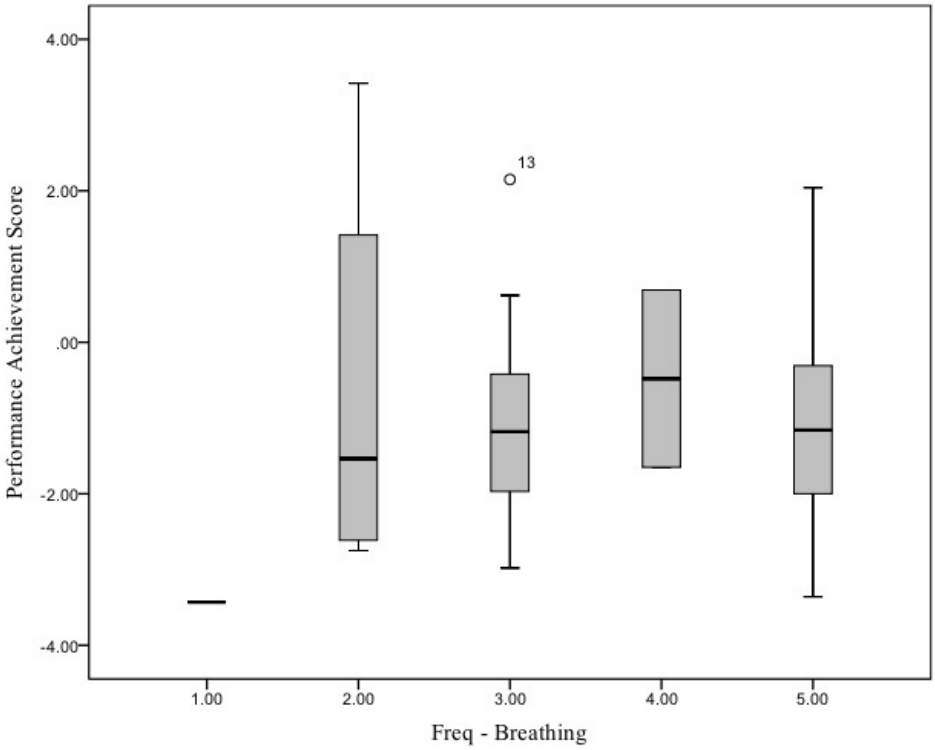

Performance Achievement Score
4.00
2.00
.00
-2.00
-4.00
15
4.00
5.00
Freq - Tuning

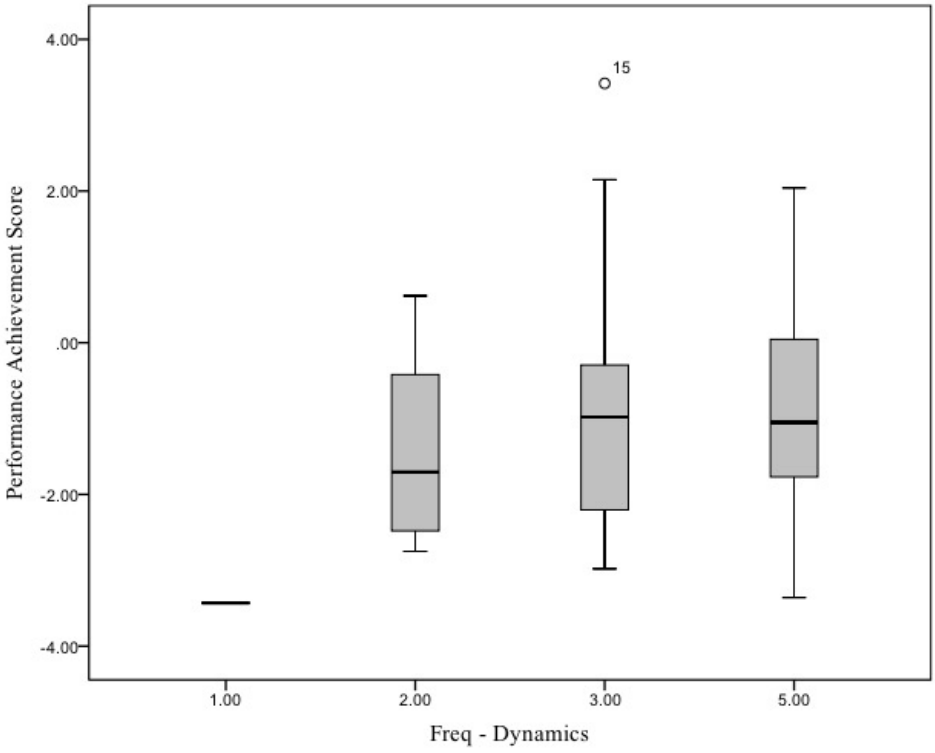

Performance Achievement Score
4.00
2.00
.00
-2.00
-4.00
13
1
1.00
2.00
3.00
4.00
5.00
Freq - Chorales

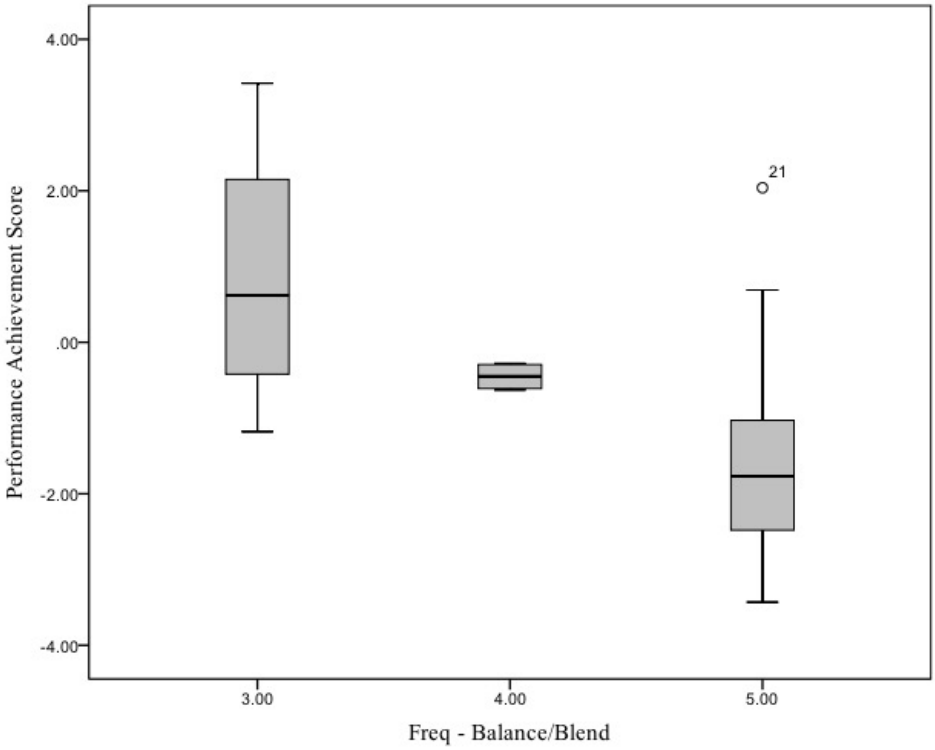

Performance Achievement Score
4.00
2.00
.00
-2.00
-4.00
15
10
1.00
2.00
3.00
4.00
5.00
Freq- Sight-Reading

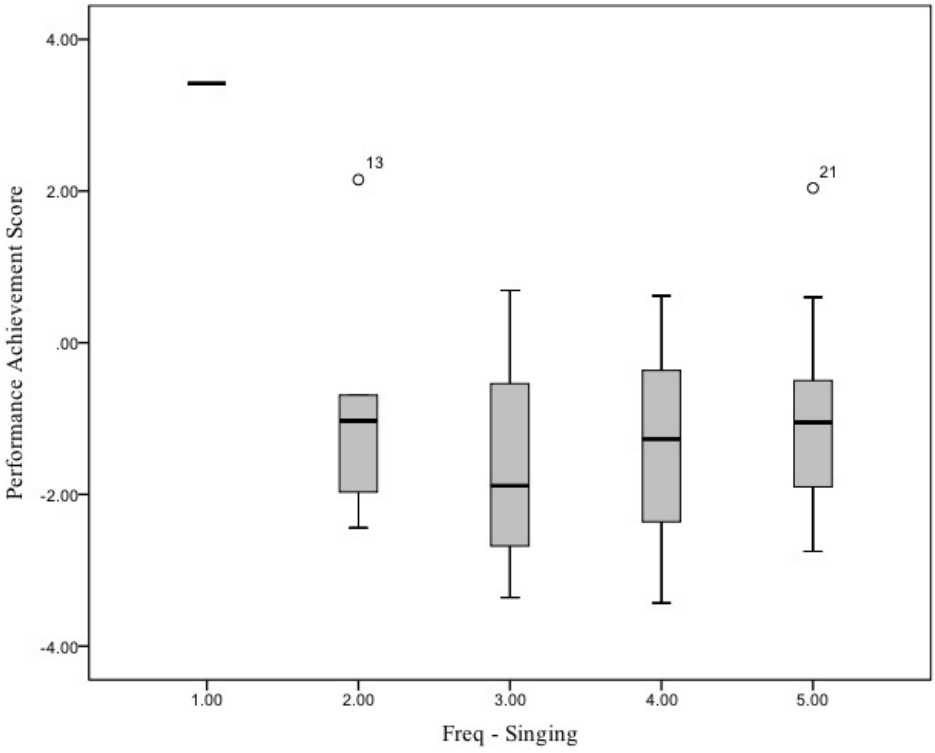

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
13
2.00
3.00
4.00
5.00
Freq - Rhythm Drills

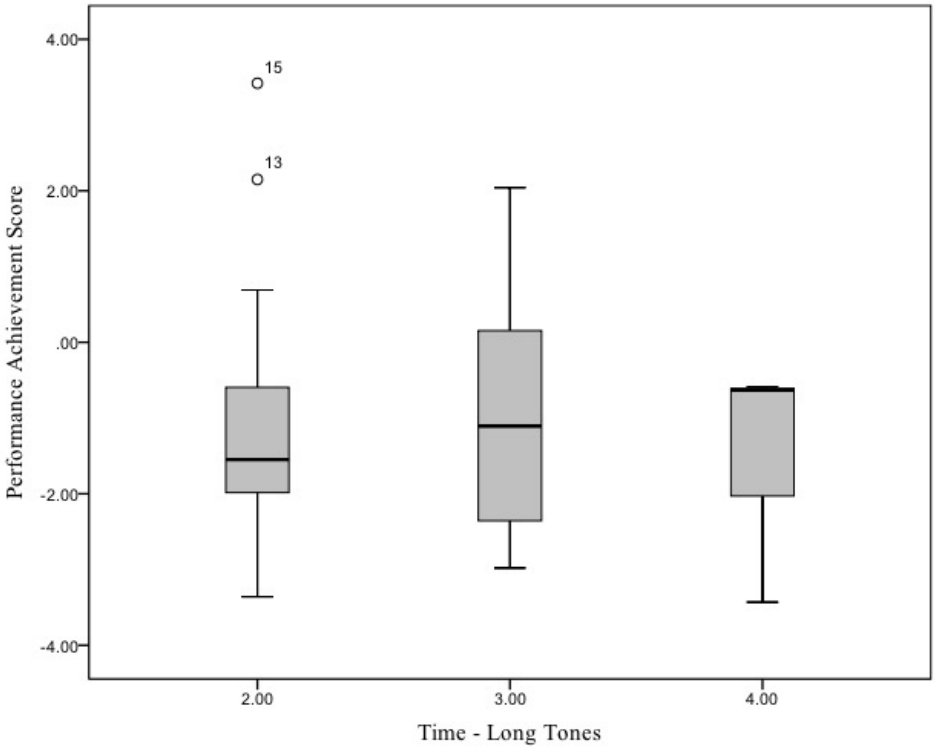

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
2.00
3.00
4.00
5.00
Freq - Articulation

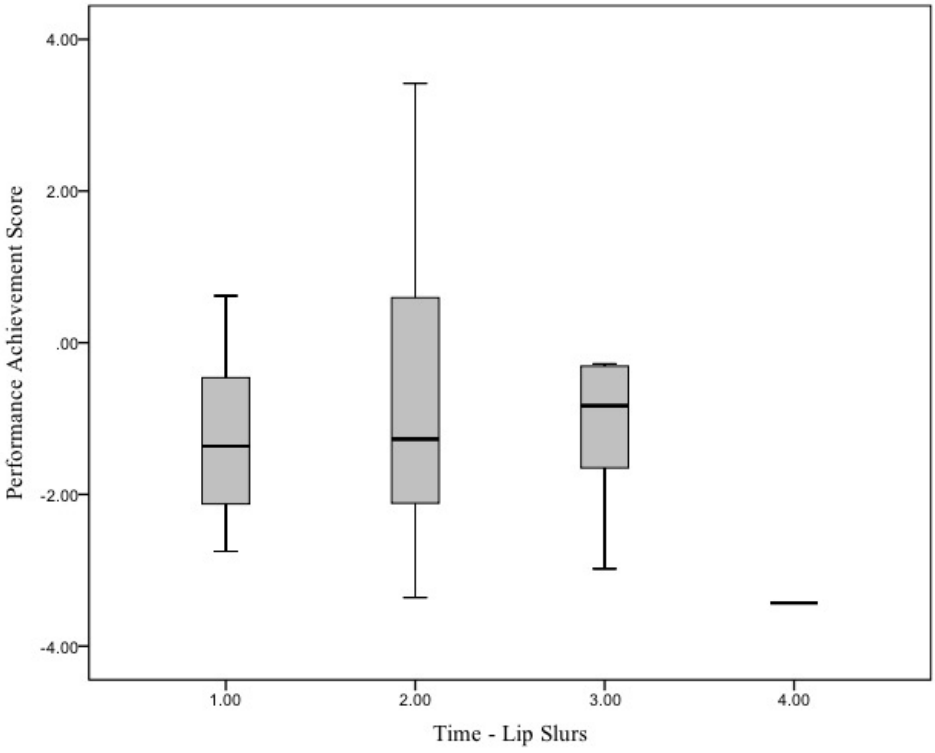

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
13
1.00
2.00
3.00
4.00
5.00
Freq - Breathing

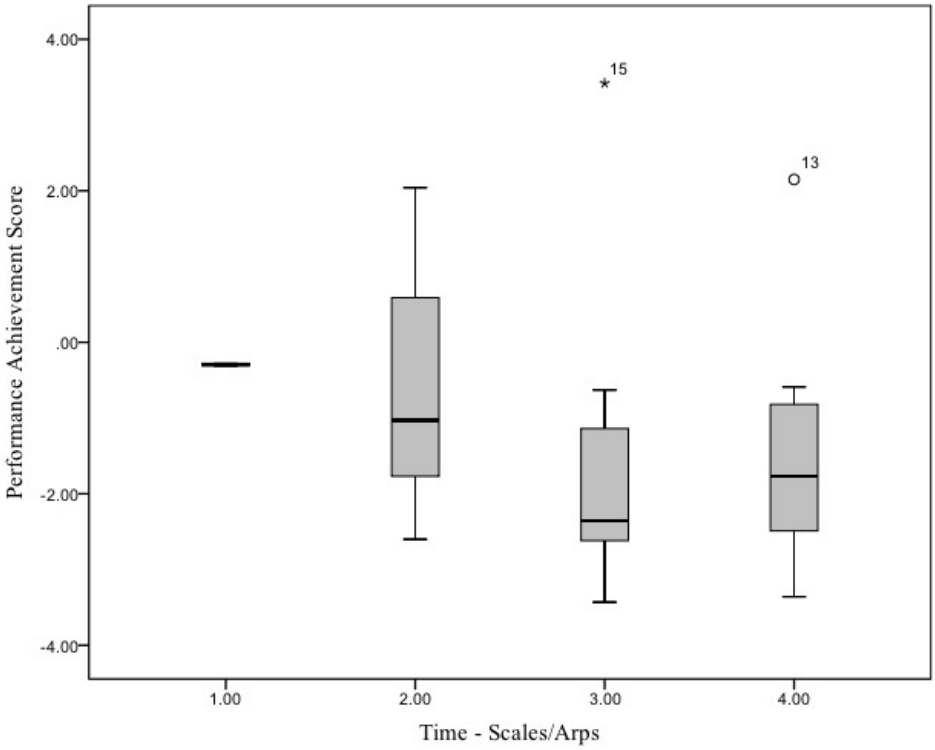

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
15
1.00
2.00
3.00
5.00
Freq - Dynamics

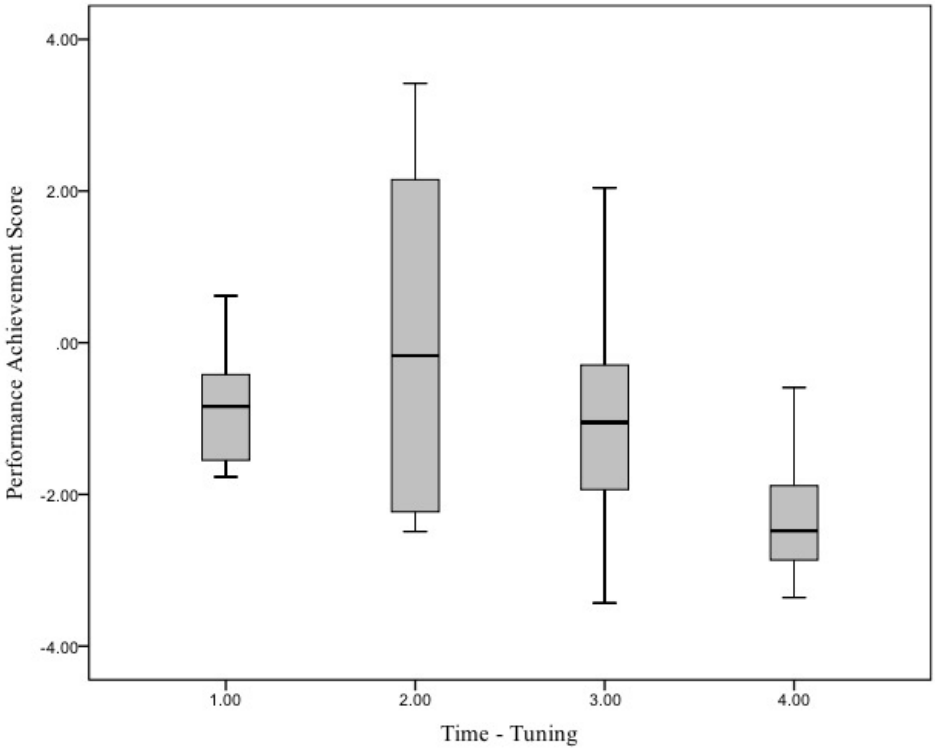

Performance Achievement Score
4.00
2.00
.00
-2.00
-4.00
21
3.00
4.00
5.00
Freq - Balance/Blend

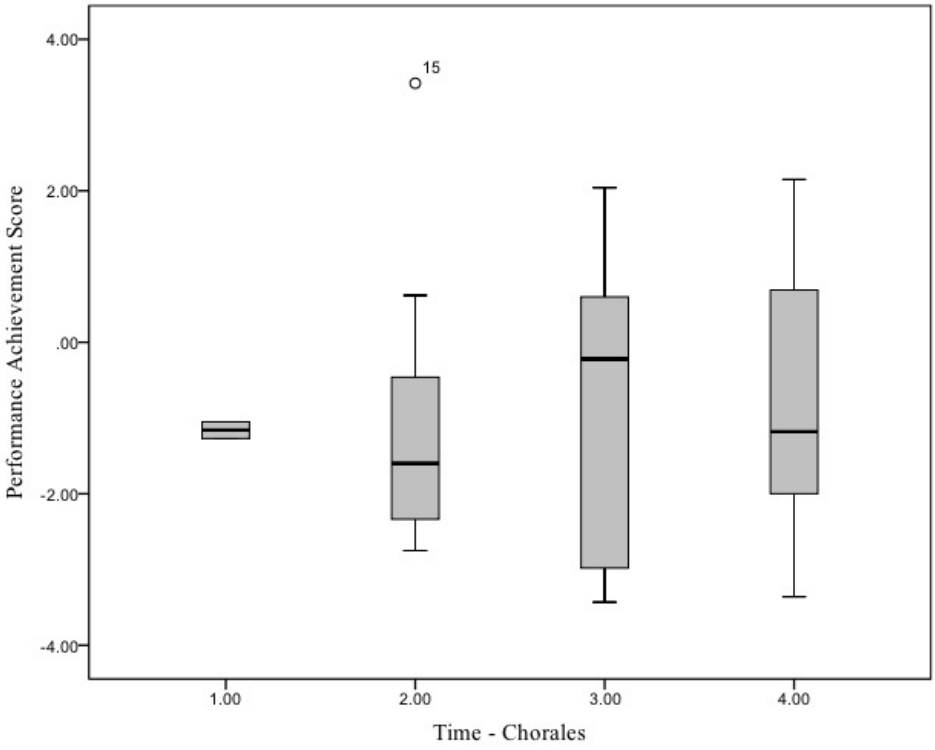

Performance Achievement Score
4.00
2.00
.00
-2.00
-4.00
13
21
1.00
2.00
3.00
4.00
5.00
Freq - Singing

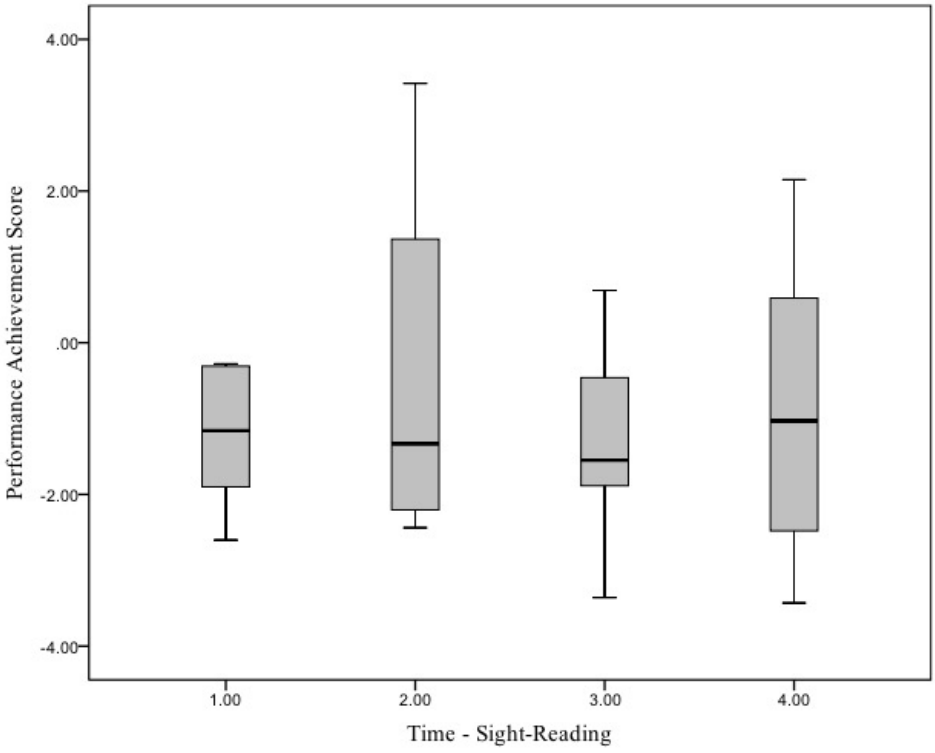

4.00
2.00
.00
-2.00
-4.00
15
13
Performance Achievement Score
2.00
3.00
4.00
Time - Long Tones

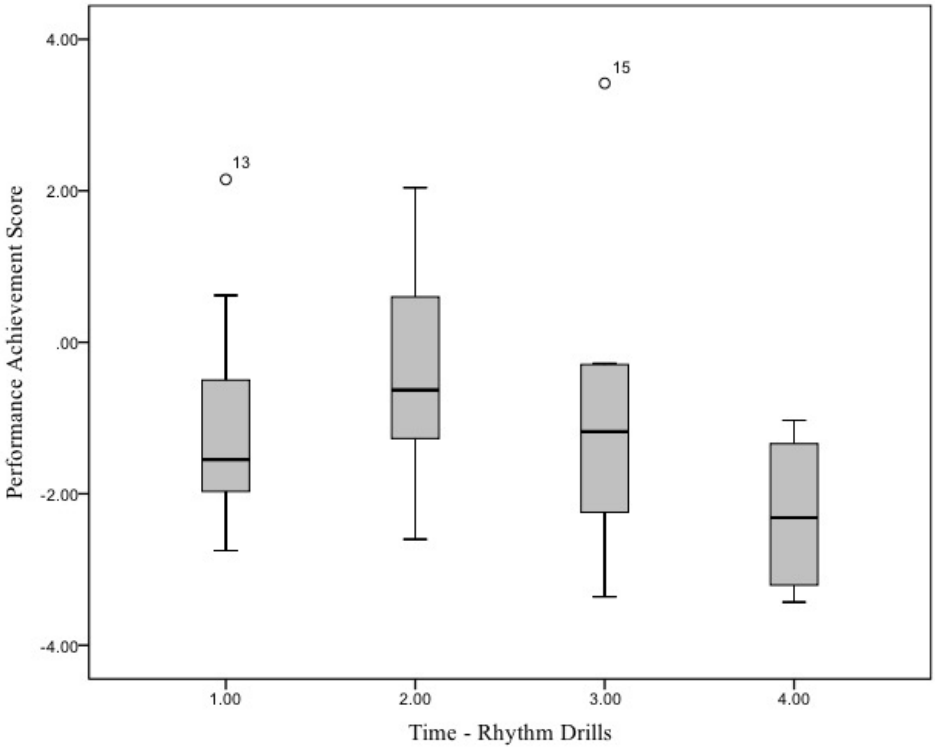

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
1.00
2.00
3.00
4.00
Time - Lip Slurs

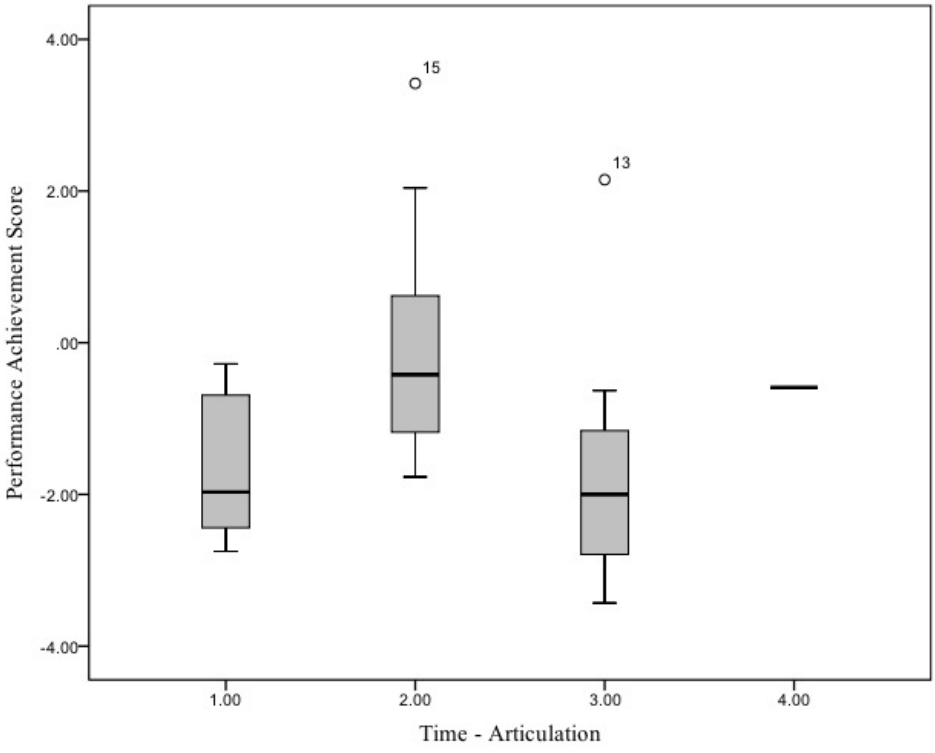

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
15
13
1.00
2.00
3.00
4.00
Time - Scales/Arps

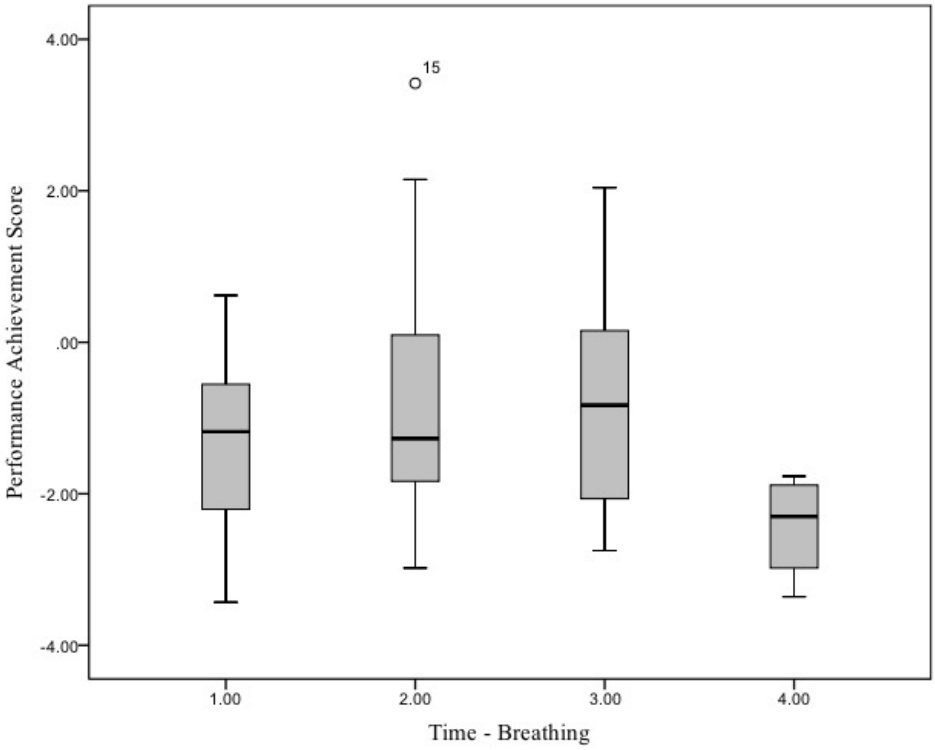

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
1.00
2.00
3.00
4.00
Time - Tuning

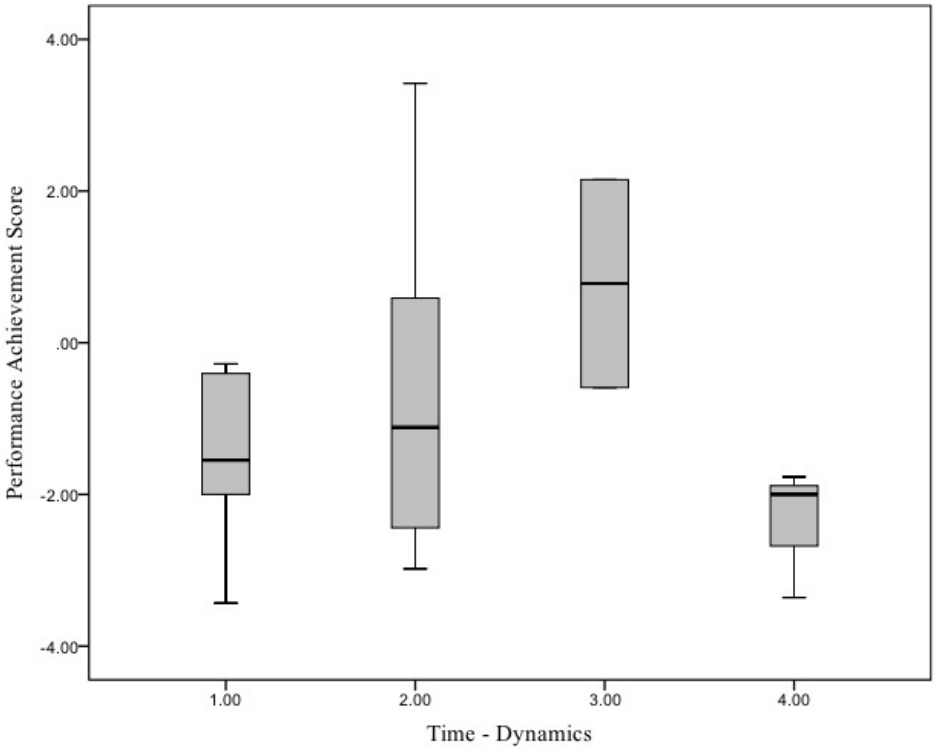

4.00
2.00
.00
-2.00
-4.00
15
Performance Achievement Score
1.00
2.00
3.00
4.00
Time - Chorales

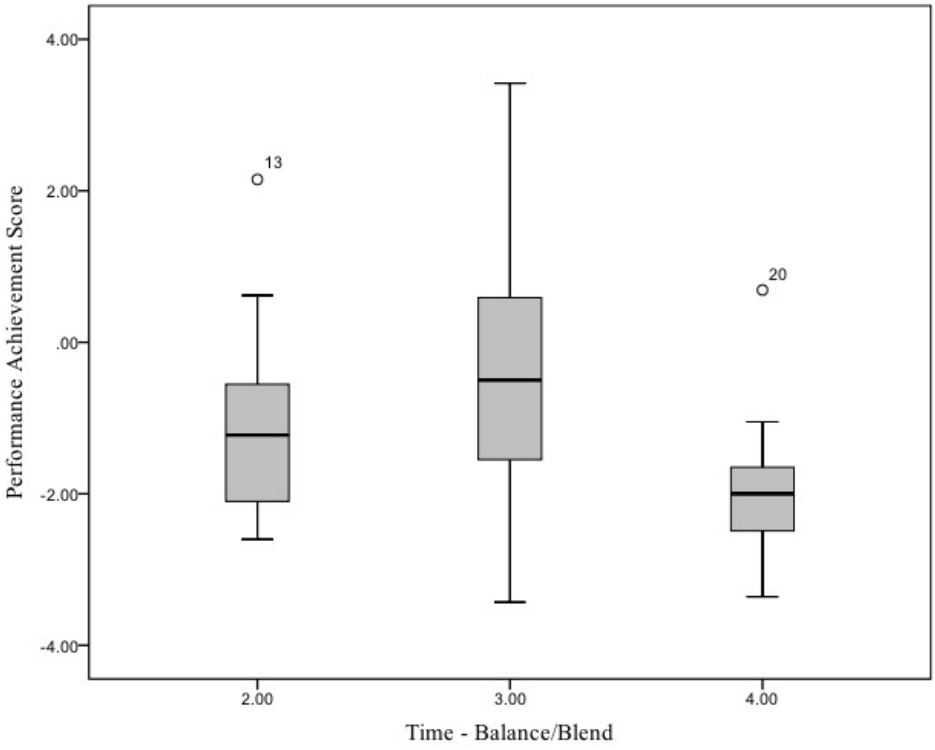

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
1.00
2.00
3.00
4.00
Time - Sight-Reading

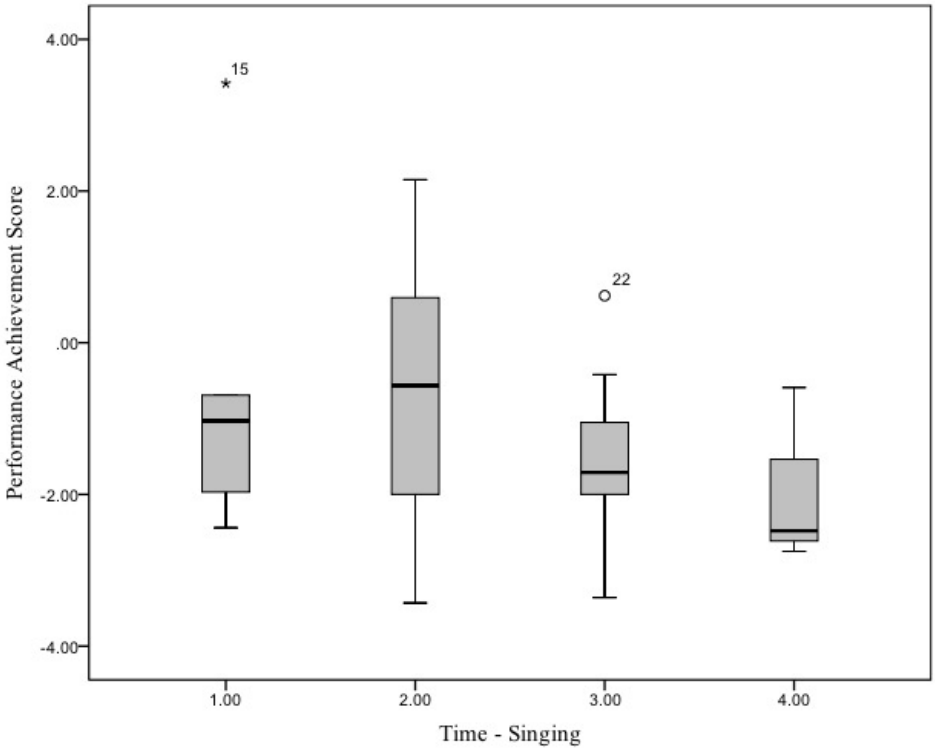

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
13
15
1.00
2.00
3.00
4.00
Time - Rhythm Drills


4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
15
13
1.00
2.00
3.00
4.00
Time - Articulation


4.00
2.00
.00
-2.00
-4.00
15
Performance Achievement Score
1.00
2.00
3.00
4.00
Time - Breathing

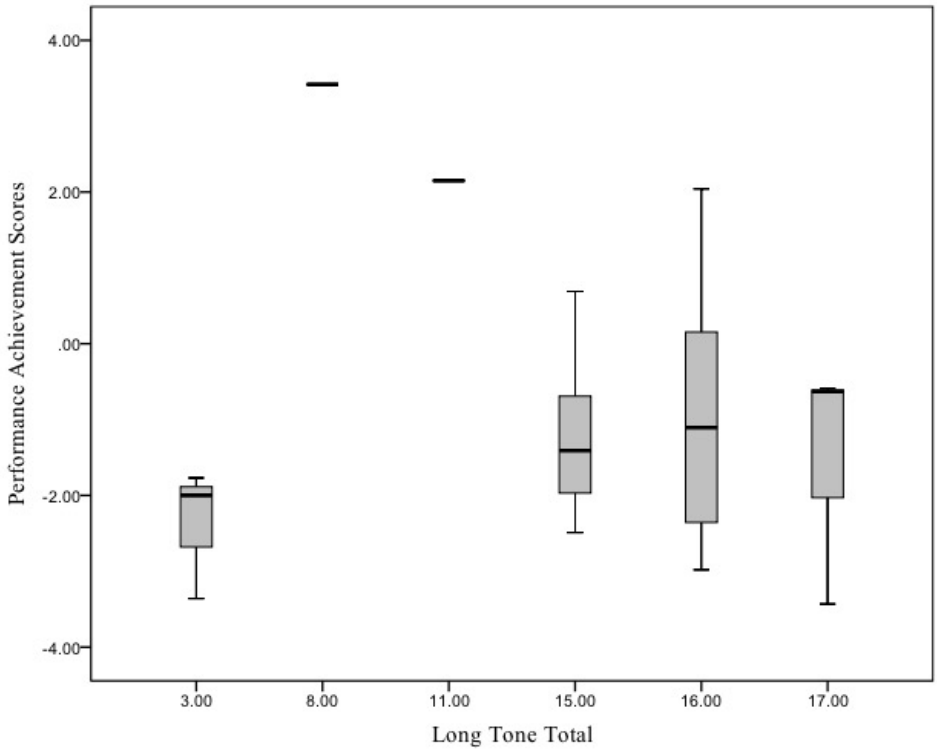

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
1.00
2.00
3.00
4.00
Time - Dynamics

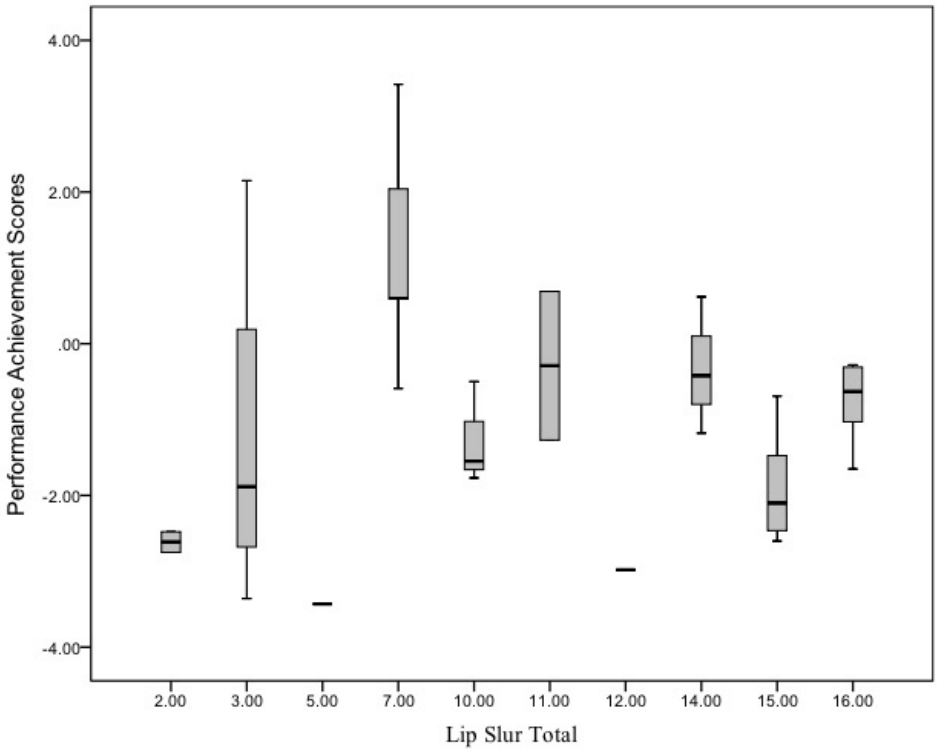

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
13
20
2.00
3.00
4.00
Time - Balance/Blend

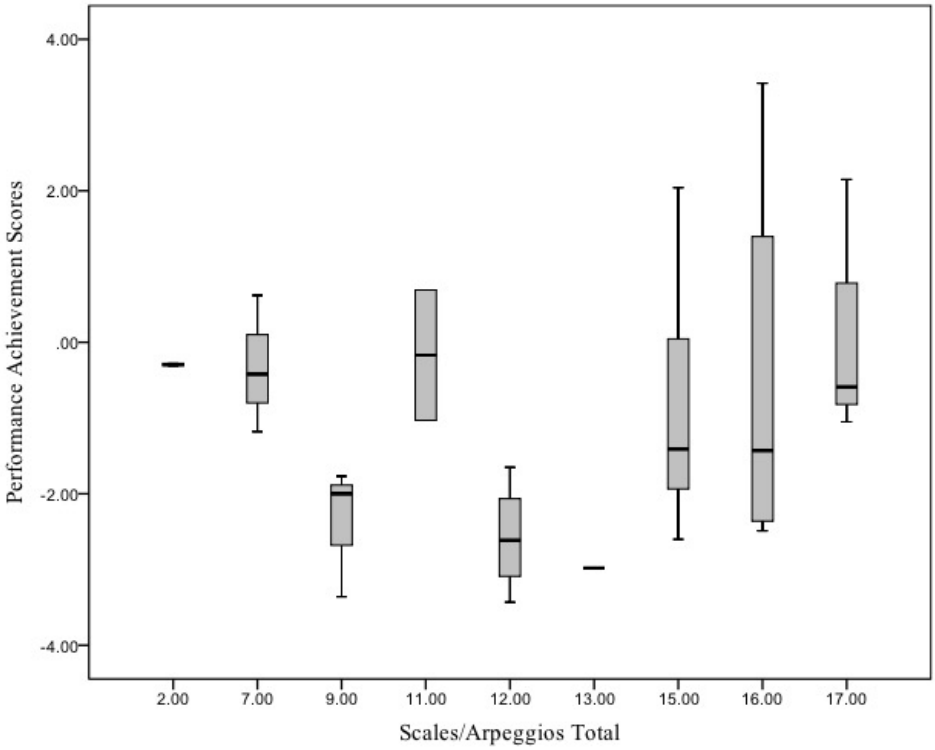

4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
15
22
1.00
2.00
3.00
4.00
Time - Singing

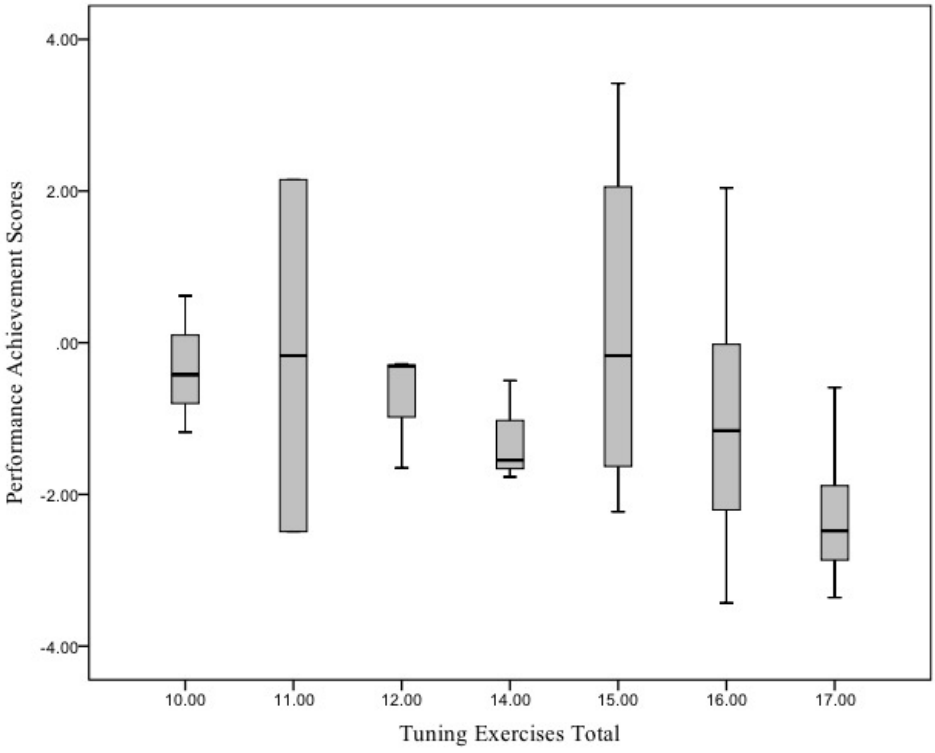
APPENDIX P

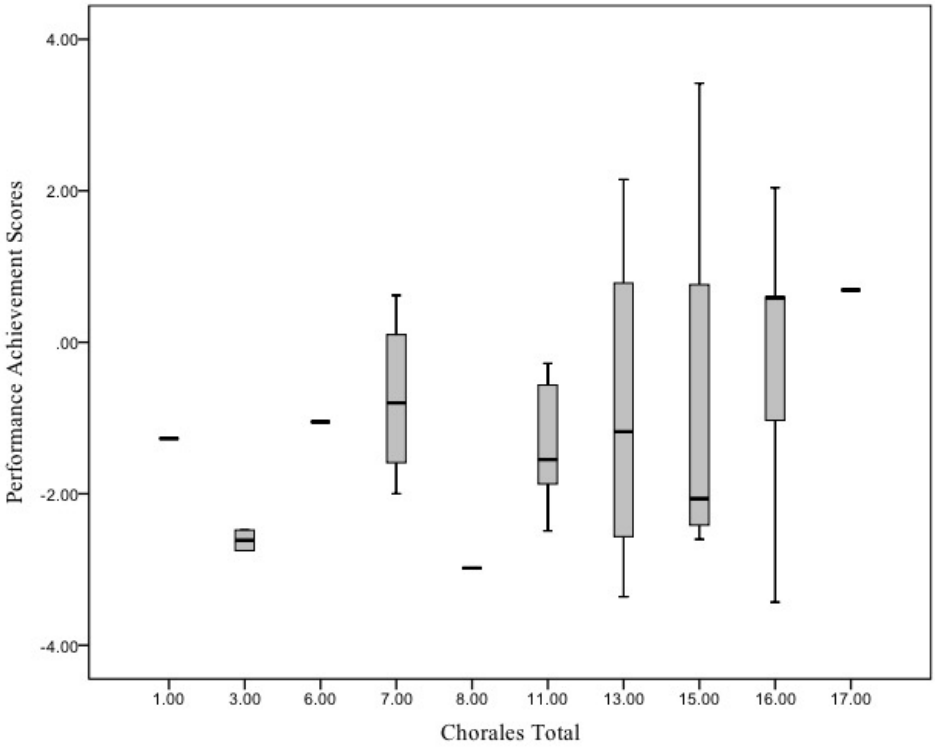
Box Plots Pertaining to Research Question 2

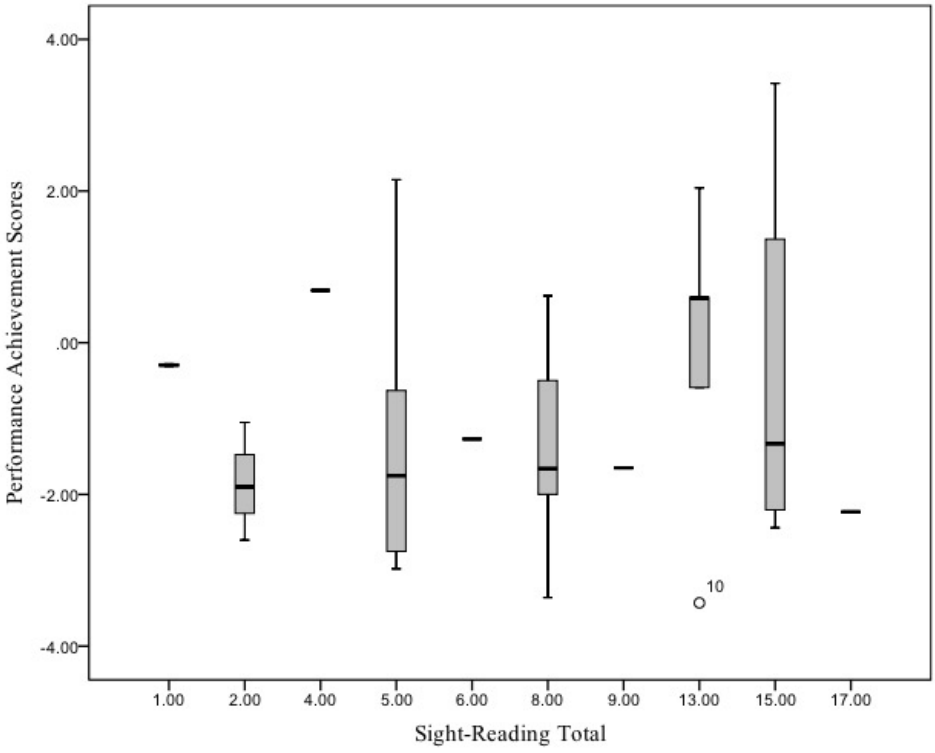


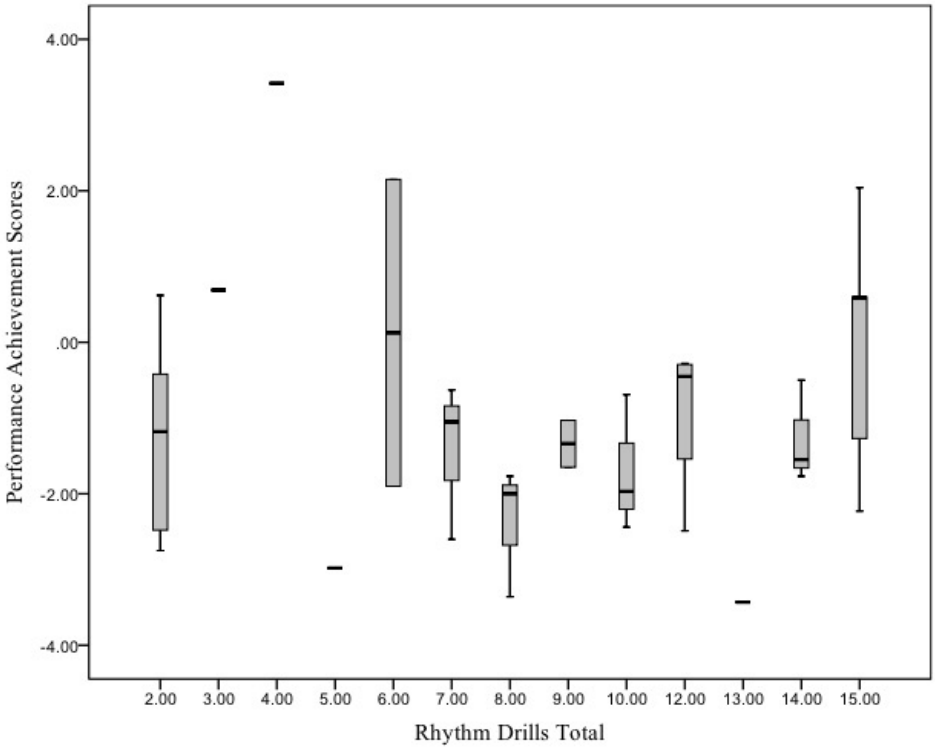


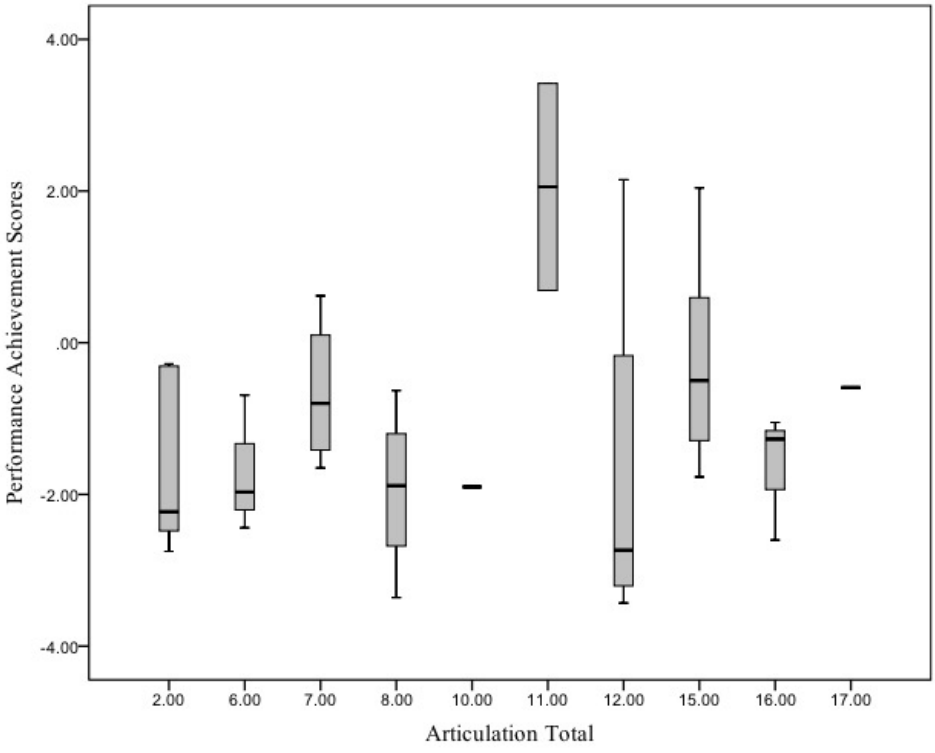

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
2.00
7.00
9.00
11.00
12.00
13.00
15.00
16.00
17.00
Scales/Arpeggios Total

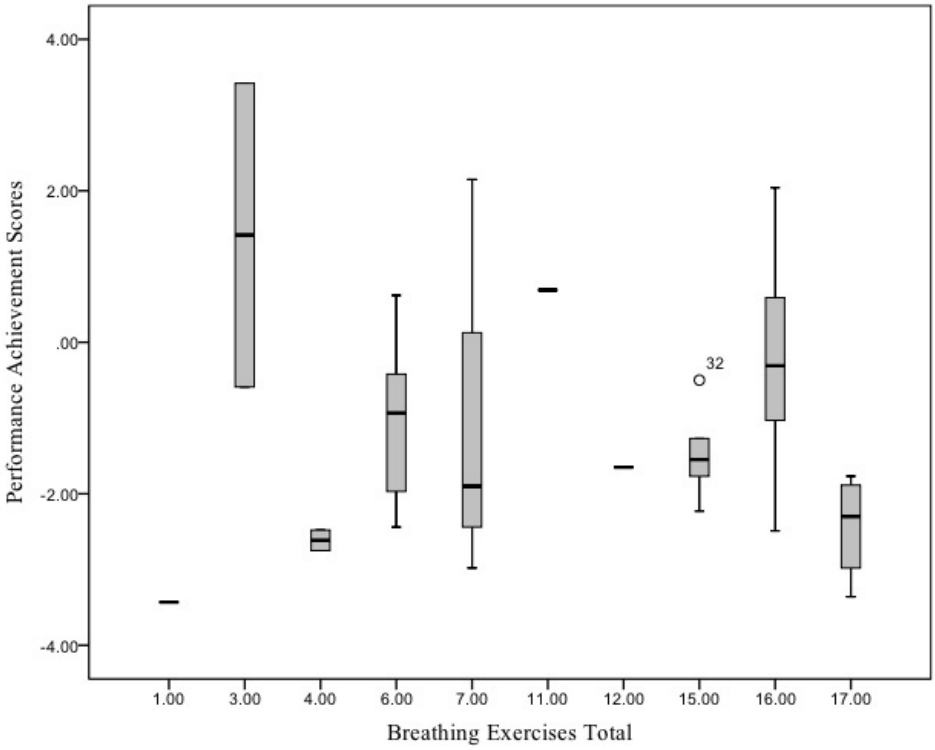

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
10.00
11.00
12.00
14.00
15.00
16.00
17.00
Tuning Exercises Total

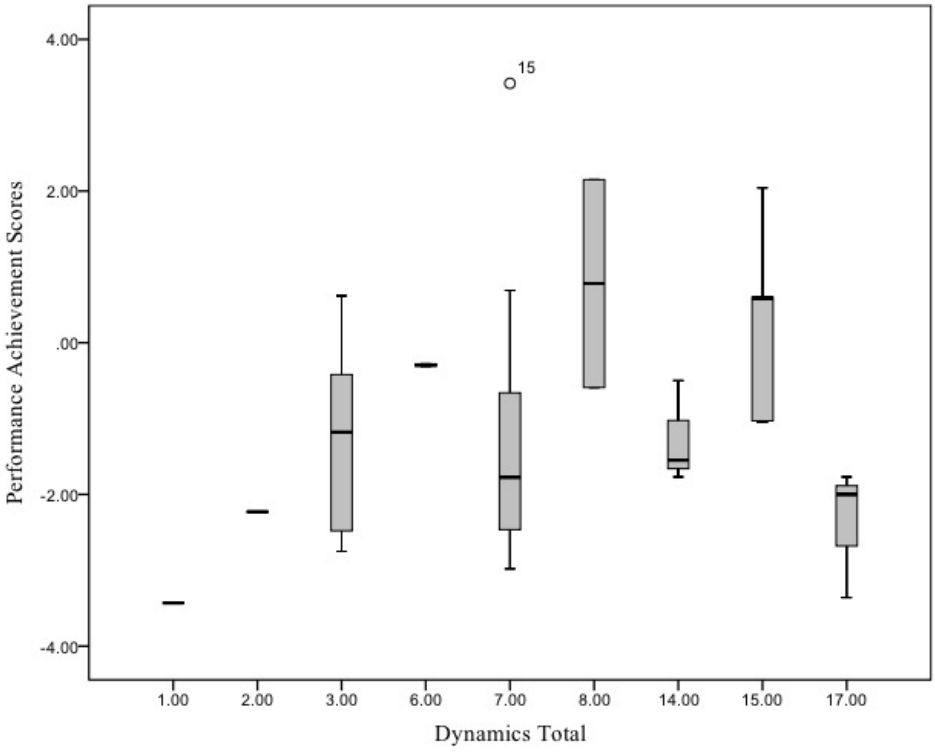

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
1.00
3.00
6.00
7.00
8.00
11.00
13.00
15.00
16.00
17.00
Chorales Total

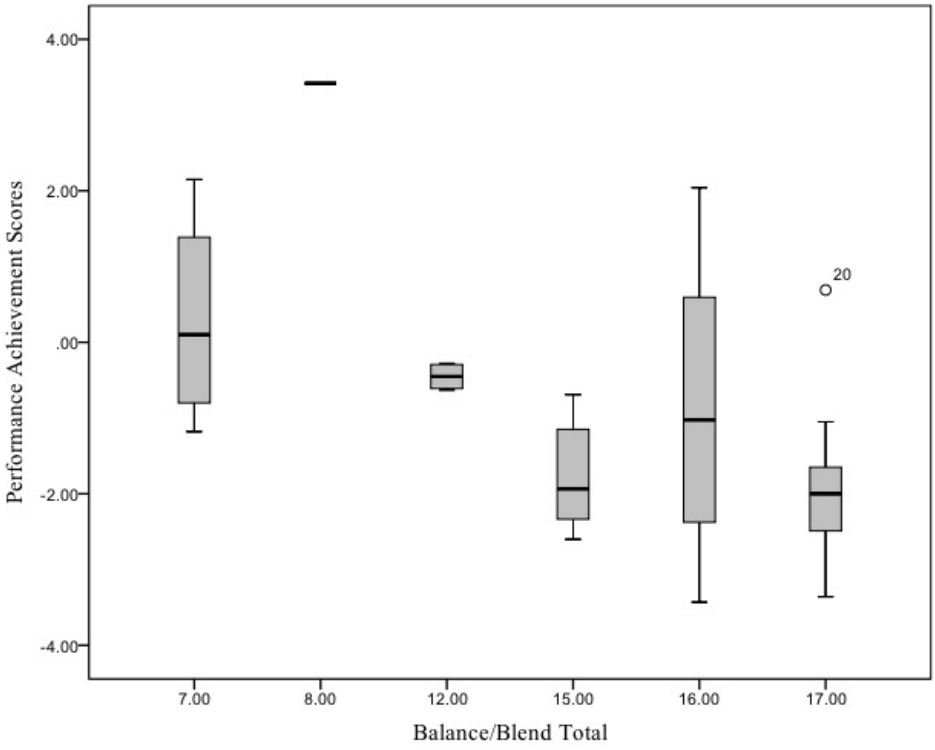

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
10
1.00
2.00
4.00
5.00
6.00
8.00
9.00
13.00
15.00
17.00
Sight-Reading Total

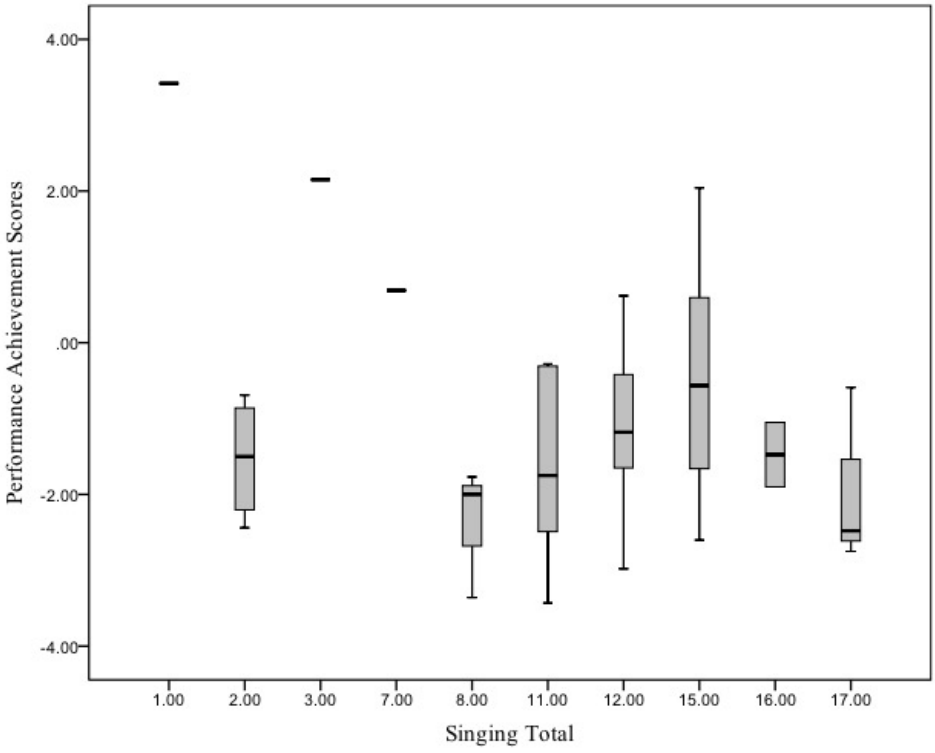

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
2.00
3.00
4.00
5.00
6.00
7.00
8.00
9.00
10.00
12.00
13.00
14.00
15.00
Rhythm Drills Total


Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
2.00
6.00
7.00
8.00
10.00
11.00
12.00
15.00
16.00
17.00
Articulation Total


Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
32
1.00
3.00
4.00
6.00
7.00
11.00
12.00
15.00
16.00
17.00
Breathing Exercises Total

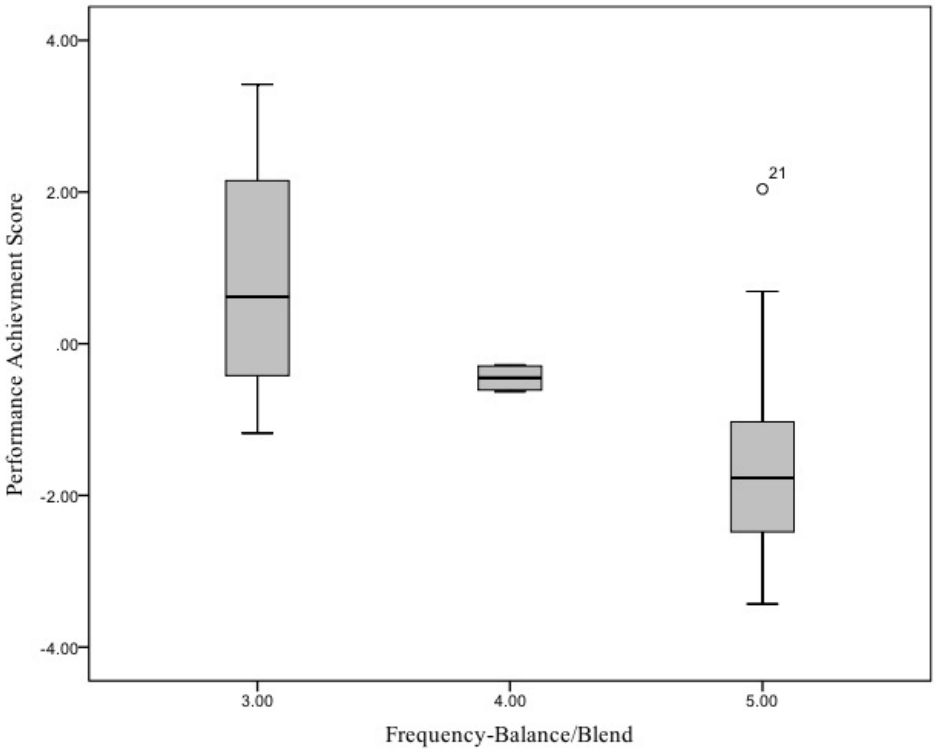

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
15
1.00
2.00
3.00
6.00
7.00
8.00
14.00
15.00
17.00
Dynamics Total

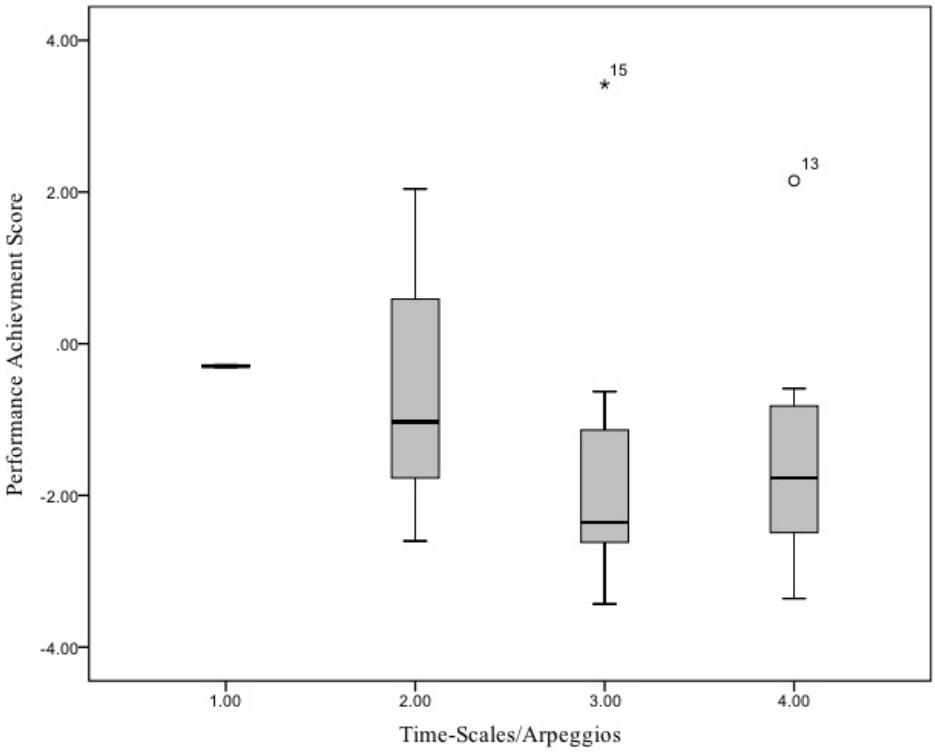

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
20
7.00
8.00
12.00
15.00
16.00
17.00
Balance/Blend Total

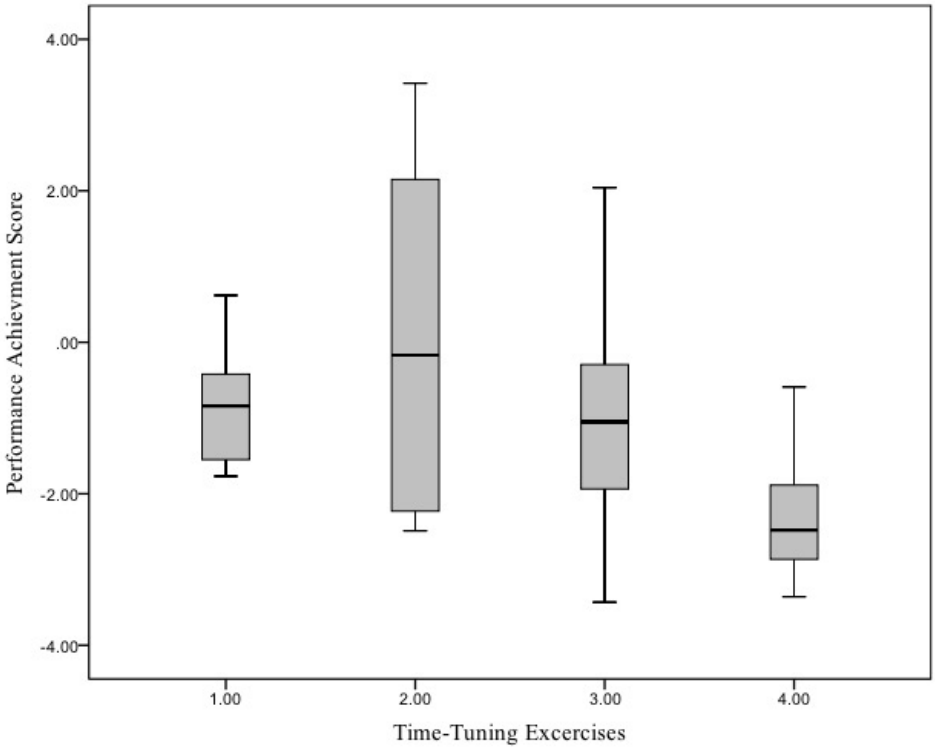

Performance Achievement Scores
4.00
2.00
.00
-2.00
-4.00
1.00
2.00
3.00
7.00
8.00
11.00
12.00
15.00
16.00
17.00
Singing Total

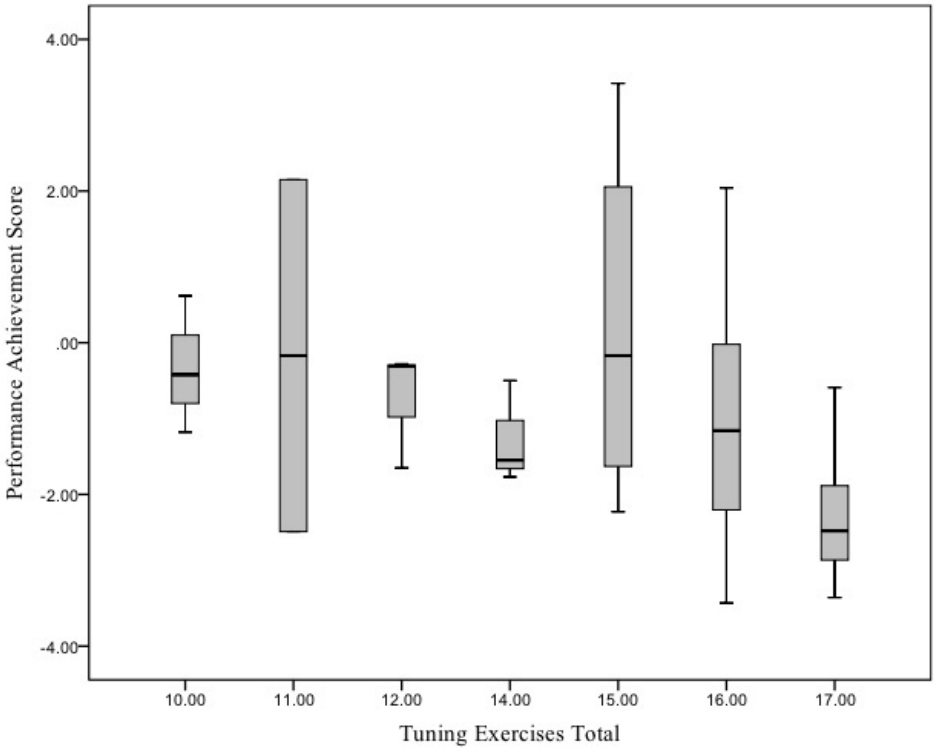
APPENDIX Q

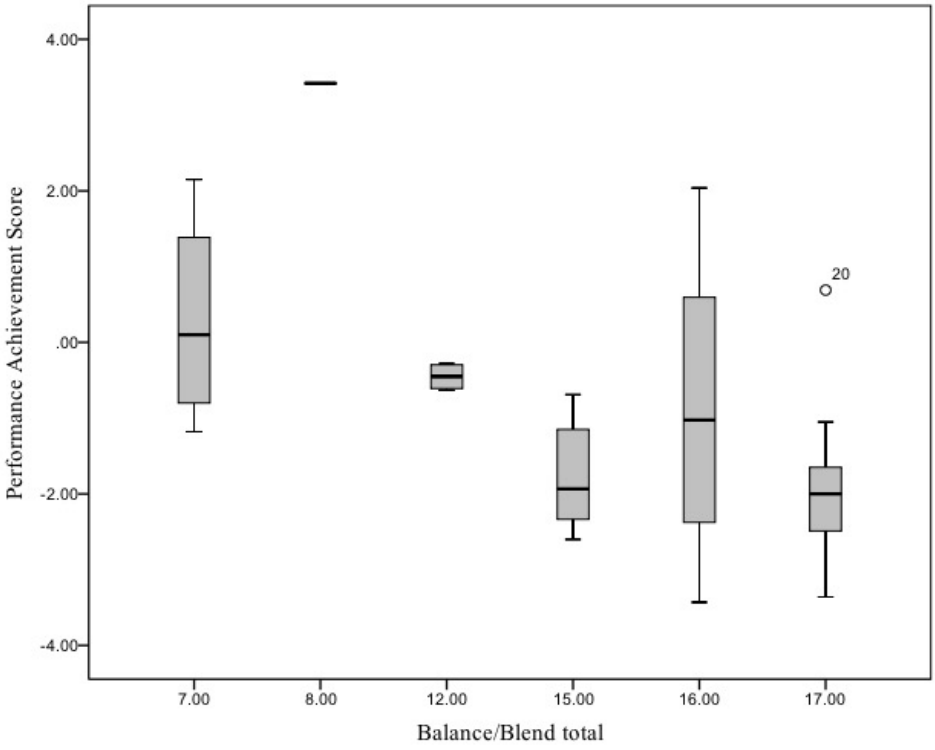
Box Plots Pertaining to Research Question 3






4.00
2.00
.00
-2.00
-4.00
Performance Achievment Score
1.00
2.00
3.00
4.00
Time-Tuning Excercises


4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
10.00
11.00
12.00
14.00
15.00
16.00
17.00
Tuning Exercises Total


4.00
2.00
.00
-2.00
-4.00
Performance Achievement Score
20
7.00
8.00
12.00
15.00
16.00
17.00
Balance/Blend total