

HANSEN, CHRISTOPHER M., Ph.D. *An Analysis of Influences on Choral Performance Adjudicators' Rating Decisions of Choral Performance.* (2017)  
Directed by Dr. David J. Teachout. 172 pp.

The purpose of this study was to (a) examine influences on choral music adjudicators' rating decisions of choral performance quality and (b) to see if differences existed among those influences by participants' years of adjudication experience and academic training (degree focus). Part One of the study included eight randomly selected participants ( $N = 8$ ), comprised of choral adjudication experts, to aid in the construction of the data collection instrument. Part Two of the study included a convenience sample ( $N = 71$ ) comprised of choral music performance adjudicators within the southeastern United States.

For the primary purpose, the *Choral Adjudicator Preference Scale* (CAPS), a researcher-constructed data-gathering instrument, was developed to determine influences on choral adjudicators' rating decisions of choral performance. A Cronbach's Coefficient Alpha measure of internal consistency was calculated to establish reliability of the CAPS data collection instrument. A coefficient of .934 was found for the CAPS, which indicated a high level of internal consistency. Validity for the data-gathering instrument was established through three sources, (a) an open-ended questionnaire sent to the eight choral adjudicators, (b) a thorough review of the related literature and (c) verification by choral activities chairpersons among the southeastern states. For research question one (What factors influence adjudicators' decisions when adjudicating choirs?) a principal component analysis revealed 23 items that coalesced among four factors of influence: (a) the ensemble's performance, (b) visual aspects, (c) extra-musical aspects, and (d) the

conductor's contributions. These four factors accounted for 61.49 percent of the total variance in participants' responses.

For the secondary purpose, a two-way repeated measures analysis of variance was calculated to determine if differences existed among factors of influence by participants' years of adjudication experience and academic training (degree focus). For research question two (Do differences exist among adjudicators' influences on rating decisions by years of adjudication experience?) a significant main effect was found for the factors,  $F(3, 189) = 216.581, p = .000, \eta^2 = .775$ ; however, there was no main effect for years of adjudication experience, nor an interaction effect among the factors and years of adjudication experience. For research question three (Do differences exist among adjudicators' influences on rating decisions by academic training?) a significant main effect was found for the factors,  $F(3, 201) = 195.326, p = .000, \eta^2 = .745$ ; however, there was no main effect for academic training, nor an interaction effect among the factors and academic training

A discussion of the influences on choral adjudicators' rating decisions was presented. Recommendations for future research were suggested regarding music performance adjudication, influences on rating decisions, and characteristics of evaluators.

AN ANALYSIS OF INFLUENCES ON CHORAL PERFORMANCE  
ADJUDICATORS' RATING DECISIONS  
OF CHORAL PERFORMANCE

by

Christopher M. Hansen

A Dissertation Submitted to  
the Faculty of The Graduate School at  
The University of North Carolina at Greensboro  
in Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

Greensboro  
2017

Approved by

David J. Teachout  
Committee Chair

© 2017 Christopher M. Hansen

This work is dedicated to the memory of the late Dr. Gary K. Wright.

APPROVAL PAGE

This dissertation, written by Christopher M. Hansen, has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair David J. Teachout

Committee Members Constance McKoy

Brett Nolker

Welborn Young

March 6, 2017  
Date of Acceptance by Committee

March 6, 2017  
Date of Final Oral Examination

## ACKNOWLEDGMENTS

The completion of this journey would have been impossible without the love and support of many people in my life. It seems impossible for me to acknowledge each individual. So, to those whose names are not listed, please know that you are in my heart and you have my deepest gratitude for helping me along the way.

Undoubtedly, this accomplishment would not have been possible without the guidance, support, and friendship of my dissertation committee members. Dr. David Teachout, there are not enough words to express my immense appreciation and gratitude for you. Thank you for being the chair of my dissertation committee. Your constant support, endless patience, and continued guidance enabled me to complete my journey. Your caring friendship inspired me along the way. You have made a profound impact on my identity as a teacher and researcher. I am honored to have had this experience of working with you. Thank you.

Dr. Brett Nolker, for years you have been a steady influence in my life. As a friend you always provided insightful advice in times of need and you have been a soundboard for my many questions, ideas, and concerns. You opened my mind to novel and creative ways of interacting with students. Your perspective continues to challenge me to grow. Thank you for your friendship and your influence on my development as a researcher, teacher, and conductor.

Dr. Constance McKoy, thank you for your heartfelt assistance and guidance throughout my doctoral work. Although I was never a student in one of your classes, our

interactions and discussions during my time at UNCG brought clarity and comfort to my work. I have great respect for you and appreciate your friendship. Thank you.

Dr. Bill Young, ever since our first meeting years ago I have admired you as a teacher, musician, and conductor. Working with you is always a joyful and memorable experience. Thank you for opening my heart and mind to being more present and responsive in the choral rehearsal and classroom. Thank you for helping me expand my abilities as an artist.

To my academic advisor, Dr. Patricia Sink, thank you for your nurturing guidance. You helped me from my first day, in every aspect of the degree program. Thank you for making yourself available and advocating for me. Thank you for helping build in me a foundation that enabled me to complete this journey.

I would feel remiss if I did not acknowledge my family. To my parents, thank you for your unconditional love. Thank you for providing me with a perspective that helped me see the world in profound ways. To my parents-in-law, thank you for the many days and nights where you provided childcare and allowed me the time and space to work. Thank you all for your love and support throughout this long journey.

My son Alexander, you were so young when we began this endeavor. Now, you are also becoming a student and scientist. I appreciated your funny episodes when you interrupted my work. Far from an annoyance, your interruptions were therapeutic—a needed break with my little guy. I hope that watching this journey taught you many lessons about life, hard work, resilience, perseverance, and most importantly, the



importance of family. Thank you for your love and support while I worked to make our lives better. I love you, son.

Last but not least, thank you to my beautiful, intelligent, and strong wife, Jamie. There are no words to describe my gratitude for having you in my life. During this journey you have been my proofreader, my editor, my cheerleader, my counsel, my rock, and so much more. This journey was not easy. But you lived this experience with me. I am so lucky to have a woman like you in my life. Thank you. I love you.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	x
LIST OF FIGURES .....	xii
CHAPTER	
I. INTRODUCTION .....	1
Background of the Problem .....	1
Philosophical Foundation for Music Performance Assessment.....	6
Need for the Study .....	11
Purpose of the Study .....	12
Research Questions.....	13
II. RELATED LITERATURE.....	14
Introduction.....	14
A Brief History of Music Festivals and Competitions.....	16
Process Model of Assessing Musical Performance .....	20
Performance Context and Purpose of the Assessment.....	23
Type of Performance.....	26
Musical Factors and Non-Musical Factors .....	29
Performance Environment .....	29
Performance Proportions .....	31
Characteristics of the Performer and the Evaluator .....	32
Appearance .....	32
Conductor's Visual Appearance .....	35
Ensemble Movement .....	36
Race Perceptions.....	37
Stage Presence and Personality.....	38
Evaluators' Personality .....	39
Evaluators' Music Experience and Ability .....	42
The Musical Performance .....	49
Training.....	50
Selection of Evaluative Criteria and Evaluative Instruments .....	55
Evaluation Criteria .....	55
Reliability.....	61
Modified Delphi Method .....	64

Summary of Related Literature.....	66
Research Questions.....	69
III. METHOD .....	70
Restatement of the Purpose.....	70
Research Questions.....	70
Participants.....	71
Data Collection Instrument.....	74
Data Collection Procedures.....	76
Data Analysis Procedures .....	78
IV. RESULTS .....	80
Review of Research Questions .....	80
Research Question 1: Factors Influencing Choral Adjudicators’ Rating Decisions .....	80
<i>Choral Adjudicator Preference Scale</i> Descriptive Statistics .....	80
<i>Choral Adjudicator Preference Scale</i> Principal Component Analysis (PCA) .....	84
Research Question 2: Years of Adjudication Experience.....	94
Research Question 3: Academic Training .....	100
Summary of Results.....	105
Participants.....	105
Methodology.....	106
Research Question 1: Factors Influencing Choral Adjudicators’ Rating Decision.....	106
Research Question 2: Years of Adjudication Experience.....	107
Research Question 3: Academic Training .....	107
V. CONCLUSIONS.....	108
Summary .....	108
Research Question 1: Factors Influencing Choral Adjudicators’ Rating Decision .....	111
Factor One—The Ensemble’s Performance .....	111
Factor Two—Visual Aspects.....	113
Factor Three—Extra-musical Aspects.....	115
Factor Four—The Conductor’s Contributions.....	117
Research Question 2: Years of Adjudication Experience.....	120
Research Question 3: Academic Training .....	121
Possible Methodological Limitations .....	122

Number of Participants .....	122
Available Data—Measure used to Collect Data .....	122
Lack of Prior Research on the Topic .....	123
Self-reported Data.....	123
Possible Limitations of the Researcher.....	125
Access .....	125
Longitudinal Effects.....	126
Suggestions for Future Research .....	126
Replicating the Study.....	127
Perception Study.....	128
Investigate Observations and Unexpected Findings from the Present Study .....	128
Conclusion .....	130
REFERENCES .....	132
APPENDIX A. OPEN-ENDED QUESTIONNAIRE TO NORTH CAROLINA ADJUDICATORS .....	147
APPENDIX B. STEP ONE: LIST OF PERFORMNCE ASPECTS PROVIDED BY EIGHT RANDOMLY SELECTED NORTH CAROLINA ADJUDICATORS .....	148
APPENDIX C. STEP TWO: COMBINED LISTS OF PERFORMANCE ASPECTS.....	150
APPENDIX D. PILOT SURVEY SENT TO PANEL OF EXPERTS.....	159
APPENDIX E. CHORAL ADJUDICATOR PREFERENCE SCALE (CAPS).....	162
APPENDIX F. INSTITUTIONAL REVIEW BOARD NOTIFICATION .....	165
APPENDIX G. RECRUITMENT LETTER TO RANDOMLY SELECTED NORTH CAROLINA CHORAL PERFORMANCE ADJUDICATORS.....	168
APPENDIX H. CAPS PARTICIPANT RECRUITMENT LETTER .....	169
APPENDIX I. CONSENT FORM.....	170

## LIST OF TABLES

	Page
Table 1. Participants' Gender .....	72
Table 2. Participants' Age .....	72
Table 3. Participants' Years of Choral Teaching Experience .....	73
Table 4. Participants' School Location .....	74
Table 5. Participants' School-Level Assignment .....	74
Table 6. <i>Choral Adjudicator Preference Scale</i> : Item Descriptions and Descriptive Statistics, Sorted by Item Mean.....	82
Table 7. Communalities.....	87
Table 8. Correlation Matrix: 23 Items .....	89
Table 9. <i>Choral Adjudicator Preference Scale</i> : Items' Individual Kaiser-Meyer- Olkin (KMO) .....	90
Table 10. Loadings and Variance Contributions for Four-Factor Structure Matrix.....	92
Table 11. Total Variance Explained: Four Factors .....	93
Table 12. Descriptive Statistics: Factors and Years of Adjudication Experience .....	95
Table 13. Participant Groupings by Years of Adjudication Experience .....	96
Table 14. Two-Way Repeated Measures ANOVA: Factors by Years of Adjudication Experience.....	99
Table 15. Pairwise Comparisons: Within-Subjects Factors .....	99
Table 16. Descriptive Statistics: Factors and Academic Training .....	101
Table 17. Participant Groups by Academic Training.....	102

Table 18. Two-Way Repeated Measures ANOVA: Factors by Academic Training.....	104
Table 19. Pairwise Comparisons: Within-Subjects Factors .....	105

## LIST OF FIGURES

	Page
Figure 1. Process Model of Assessing Musical Performance.....	22
Figure 2. Scree Plot for <i>Choral Adjudicator Preference Scale</i> Responses.....	86

## CHAPTER I

### INTRODUCTION

#### **Background of the Problem**

Imagine you are a student member of a high school choir. Your ensemble is preparing a performance for the state's music contest. Your choral director selected two or three songs to be performed by your ensemble at the event, and for many weeks and months your choir routinely rehearsed in preparation for the music contest. As the event approached, your choir polished the musical performance in every possible way. Finally, the day arrived. Your music was memorized. You worked on your stage presence. You practiced how to walk on and off the stage. Your choir uniform was cleaned and pressed. It was time for your ensemble to perform.

Now, imagine that you are one of the adjudicators at a high school choral music contest. Your job, as a performance judge, is to observe choirs' performances and assign each a quality rating. You have years of teaching experience, choral conducting experience, and maybe an advanced degree in choral music education and conducting. Prior to the music contest you received copies of the event's ratings forms and rubrics used for evaluating the choirs. You review the forms and feel that you have a clear understanding of what is expected of you as a performance judge. Finally, the day of the event has arrived. It is time for you to adjudicate the choral performance contest.



The scenario described above provides the setting of a choral music performance contest, in which a choir is about to perform for a panel of adjudicators and receive a performance quality rating. If the scenario had continued to play out what might have happened? What ratings might the adjudicators have assigned to the choir's performance? All possible outcomes from this scenario would depend on the choir's performance and the interplay between a myriad of complex physical and mental operations of the adjudicator (Landy & Farr, 1980).

In some cases, the choir might have received a similar rating from each adjudicator, however; choral directors indicated that there were times when adjudicators' ratings were too severe or too lenient, were insufficient regarding comments or feedback, or had too much variance among adjudicators' ratings (Fox, 1990). Payne (1997) stated, "The reliability [level of agreement among judges] of festival results is a continuing source of controversy" (p. 8). This statement suggested level of agreement among adjudicators at music contests might not always be acceptable. Payne (1997) summarized Fisk saying, "using a sample of almost five hundred musicians ... even experienced adjudicators exhibited a dangerously low reliability coefficient for their festival ratings" (p. 8). What might contribute to unacceptable reliability among adjudicators? Researchers suggest the cause is due to bias (Ekholm, 2000; Ford, 1999). Therein lies the problem: Music educators and researchers have expressed concerns regarding adjudicator bias and its possible influence on rating decisions during music performance.

Ekholm (2000) concluded that adjudicators' preferences influence and inform their rating decision regarding music performance. Participants with solo-voice expertise preferred a soloistic choral tone while choral conductors preferred a blended choral tone (Ekholm, 2000). From this result, one might conclude that choral performance adjudicators with a solo-voice focus might be inclined to rate tone quality differently than choral conductors. Adjudicators with formal vocal training, but are not choral conductors, might contribute to inconsistencies among adjudicators' scores due to their bias regarding desired vocal tone qualities.

Researchers investigated aspects regarding characteristics of the performer and evaluator and how those characteristics might influence perceptions of music performance quality. These studies focused on aspects about participants' personality traits (Thomas, Silvia, Nusbaum, Beaty, & Hodges, 2015), music experience and ability (Tsay, 2013), as well as gender and race (Cheek, 2007; Elliot, 1995/1996). Although participants claimed that their judgments [rating] regarding music performance quality were not influenced by visual aspects of the performance, researchers offer evidence that peoples' perceptions of choral music performance quality are determined by an interaction among the senses (Mitchell & MacDonald, 2016; Tsay, 2013).

There have been ongoing discussions among researchers and practitioners regarding the influence of the performance environment and non-musical aspects on adjudicators' ratings of music performance (James, Demaree, & Wolf, 1984; McPherson & Thompson, 1998; Radocy, 1989; Robinson, 1990). In general, results from prior research studies support the idea that adjudicators' perceptions of musical performance

quality are influenced by time, place, and environment. Adjudicators' ability to concentrate their attention seemed to be affected by variables including musical score use (Napoles, 2009), fluctuations in mood (James et al., 1984; McPherson & Thompson, 1998), time of day (Bergee & Platt, 2003), room conditions (Daugherty, 1999), visual aspects of the ensemble (Howard, 2012), visual aspects of the conductor (Silvey & Risher, 2015), and personal characteristics (Lehmann, 1997). Are adjudicators aware of how they might be affected by personal and contextual factors? Although research supports the notion that environmental and non-musical aspects influence adjudicators' perceptions of performance, how might a self-report design about influences on adjudicators' rating decisions of choral music performance quality affect study results?

Do differences exist among adjudicators regarding influences on their rating decisions of choral performance by years of adjudication experience? Available research supports the notion that peoples' experiences with music and teaching might influence how they listen to and evaluate music; however, there is limited research regarding the variable of years of adjudication experience. Prior research focused primarily in two areas: (a) instruction in music classes as experience (Hair, 1981) and (b) teaching-career level (years teaching) as experience (Hewitt & Smith, 204). One researcher mentioned 'adjudication experience' as training for music performance assessment (Mills, 1991); however, the aspect of adjudication experience was not the focus of the study. Logically, years of adjudication experience would seem to influence how an adjudicator rates music performance.

Researchers discovered and discussed inconsistencies among adjudicators' numerical ratings at both the state and national levels across a variety of high school music performance event types (Bergee, 2003; Cooksey, 1977; Corbin, 1995; Fox, 1990; Hash, 2012; Latimer, 2007; Latimer, Bergee, & Cohen, 2010; Norris & Borst, 2007; Robinson, 1990; Ward, 2013). Some researchers found evidence for concern regarding reliability [level of agreement] among adjudicator's evaluations of choral performance (Cooksey, 1977; Latimer, 2007; Latimer et al., 2010). The most reoccurring problem found in these studies was related to the inter-rater reliability among evaluators (Bergee, 2003; Hash, 2012; Latimer, 2007). Other researchers have conducted studies on the selection of evaluative criteria and evaluation (Jones, 1986; Wrigley & Emmerson, 2011) with the intention to improve the music performance evaluation process.

How might training influence adjudicators' performance at music performance adjudication? Many researchers referred to training in specific areas, for example: general music training (Hair, 1981; Juchniewicz, 2008; Tsay, 2013), vocal or instrumental training (Ford, 1999), experience as training (Mills, 1991; Kinney, 2009), and training in adjudication (Ward, 2013). Adjudicators represent a diverse background of musical training, education, and experience. They might hold one, two, or three degrees. Some might have specialized certificates in one or more domains. Some degree types are education-focused, performance-focused, or a hybrid of education and performance specialties. No studies were found in which adjudicators' academic degree and focus as training was investigated. Because many state-level music education associations hire adjudicators from colleges and universities, primary and secondary

schools, and private music businesses, investigating possible differences among academic training may reveal important information.

What factors might influence adjudicators' evaluations of choral performance? McPherson and Thompson (1998) stated, "Assessing musical performance is common across many types of music education practices, yet research clarifying the range of factors which impact on a judge's assessment is relatively scarce" (p. 12). Researchers investigated various aspects of music performance adjudication and determined that the performance environment (McPherson & Thompson, 1998), characteristics of the adjudicator (Thomas et al., 2015), visual aspects of the conductor and performing ensemble (Madsen, 2009; Mitchell & MacDonald, 2016), as well as musical aspects (Cooksey, 1977), influenced adjudicators' perceptions of musical performance. The present researcher was unable to locate research regarding self-reported influences on choral performance adjudicators' rating decisions of choral performance, years of adjudication experience, and academic training. For the current research study, rather than examine the evaluation instrument, the performance environment, or the process, the present research study investigated influences on choral performance adjudicator's ratings of choral performance to better understand what aspects influence adjudicators' decisions regarding ratings.

### **Philosophical Foundation for Music Performance Assessment**

Researchers agree that state music education associations, including music education policy makers and event organizers, should have a clear philosophical foundation from which to guide their music performance assessment practices (Batey,

2007; Corbin, 1995; Fox, 1990; Stegman, 2009). Batey (2007) suggested that the most important factor concerning music performance assessment is to provide young singers (students) positive educational experiences. This suggestion might possibly provide insight into a philosophical foundation for music performance assessment—to provide singers with a positive educational experience.

Corbin (1995) discussed music adjudication situations. The actual music performance adjudication event can be a place for choirs to hear other choirs. Choral directors and choirs can share among their peers in other ensembles and among many schools. Furthermore, Corbin suggested the concept of the ‘final product’ might be different for each choir dependent on contextual characteristics of the choir. This concept of the final product might be considered a structural pillar, or imperative component, of a philosophical foundation of a music evaluation event.

For many adjudicators and ensembles, the festival or competition rating seemed to be the ultimate measure of an ensemble’s performance quality (Corbin, 1995; Freer, 2011). Unfortunately, the results from such an event might be detrimental. In many instances, especially in education, the school principal might expect the choir to receive high rankings or be victorious over other ensembles. This expectation might then be passed down from the principal to the director and held over the singers (students). According to Corbin, “this [assumption of rating importance] puts unnecessary pressure on everyone and encourages deviation from quality teaching” (p. 49).

Fox (1990) believed a cause of concern with music performance adjudication is the philosophy of the event. Fox suggested that music performance adjudication event

organizers, music educators, and adjudicators articulate clearly the event's goals and ensure those goals are reflected on the measurement instrument—rubric or ratings form. Fox (1990) also suggested music performance adjudication event goals and measurement instruments should be printed and disseminated to music educators and adjudicators prior to the event. Additionally, adjudicators should meet prior to the festival to discuss the goals of the event and review the measurement instrument.

Stegman (2009) provided an exemplar of a state music educator association's complete revision of a music performance adjudication philosophical framework, measurement instruments, and practices. The author examined the Michigan School Vocal Music Association's (MSVMA) music performance evaluation. The MSVMA suggested the primary focus of music performance evaluation was, "improving, promoting, and assisting vocal music education in the middle, junior high, and senior high schools of the state through a program of clinics, festivals, workshops, contests, and conferences, regulated by legitimate and effective means" (p. 63). The organization created an adjudication philosophy that governed the philosophical framework over all vocal programs. The organization created a new music performance measurement instrument (evaluation sheet) and redesigned their adjudicator recruitment and training process and procedures.

South Carolina Music Educators Association (SCMEA) described the purpose of music performance adjudication as an opportunity for a singer or ensemble to experience adjudication (South Carolina Music Educators Association, 2015). As stated in the organization's 2015 to 2016 Choral Division Handbook, "The Solo and Ensemble festival

provides students the opportunity to perform before an adjudicator. The adjudicator provides a written and verbal critique of the performance and awards a rating based on defined criteria” (South Carolina Music Educators Association, 2015, p. 15). This was the only statement within SCMEA’s governing documents that provided a rationale for choral music performance adjudication.

The Alabama Vocal Association (2015), a division of the Alabama Music Educators Association (AMEA) stated, “State Choral Performance Assessment provides school choirs with a venue for adjudicated public performance of quality repertoire and closed sight-reading skills assessment” (p. 11). This statement does not define the term ‘quality repertoire’ nor does it elaborate on possible deeper reasons for having such an event, however; it did provide a rationale for the event. The Alabama Vocal Association’s Bylaws provided a thorough description of policies and procedures pertaining to their choral performance assessment including, but not limited to (a) assessment will occur in each district, (b) choirs must have a minimum of 17 members, and (c) choirs will perform two songs from memory (Alabama Vocal Association, 2015).

Florida School Music Association (2015) described the purpose for music performance assessment in the state of Florida:

An MPA is a major assessment event for secondary music programs in which a musical performance is evaluated on specific criteria. The assessments are held annually for marching bands, concert large ensembles (band, chorus, orchestra), and solos-ensembles. Together with other criteria...the music performance assessments may give a significant measure of success of a music program on a campus. (p. 3)



The Florida School Music Association (FSMA) is responsible for statewide music assessment and reports to the state's District School Superintendents, School Principals, Music Supervisors, and District Contacts. The FSMA President writes an annual report detailing all school music programs ratings and statistics that compare schools and districts. Furthermore, the FSMA provides brief descriptors for ratings (Florida School Music Association, 2015).

Choral performance adjudication in North Carolina claims to focus on the improvement of a choir's performance abilities while providing an evaluation based on standardized criteria and on comparisons to other choirs (North Carolina Music Educators Association, March 2015). As stated in the organization's policies and procedures, the purpose of music performance assessment is, "to provide each performing group with an opportunity to improve and to evaluate its performance by comparison with a set standard of excellence, and by comparison with the performance of other groups" (North Carolina Music Educators Association, March 2015). This statement is the only rationale for why and how music performance adjudications occur and function.

Researchers agree on the importance for music educator associations to develop and clearly articulate a philosophy that acts as the primary focus for music performance evaluation (Fox, 1990; Stegman, 2009). With increased emphasis on standardized testing, event organizers, adjudicators, and music educators face an increasing responsibility to develop and agree on a clear philosophy for music performance assessment.

Adjudicators can use the philosophy of an event to help them understand better the expectation for ensembles and soloists at various secondary levels. Music educators (directors, conductors, and private vocal instructors) can use the information presented by adjudicators as a guide for successful ensemble and soloist preparation for music performance adjudication.

### **Need for the Study**

A general understanding among high school chorus teachers and program stakeholders is that music performance evaluations are intended to be positive educational experiences for both students and teachers. State-level music educator association administrators, school-based administrators, and chorus teachers see state music performance adjudication as synonymous with a state standardized test, using the rating as a means to represent the quality of teaching and learning by comparing the rating to the approved standard (Florida School Music Educators, 2015; Freer, 2011; Guegold, 1989; Rohrer, 2002). Rohrer (2002) stated ratings at music competition “became a unit of measure for many communities and school systems for evaluating music programs and directors” (p. 41). Forbes (1994) stated, “Today, with the increased emphasis on accountability in education and the demand for evidence of student achievement, the use of contest and festival ratings as indicators of teacher effectiveness seems more prevalent than ever” (p. 16). For many ensembles, and music directors, the festival or competition rating is perceived to be the ultimate measure of an ensemble’s performance quality (Corbin, 1995). Local education agencies, school administrators, choral directors, students, and surrounding communities might use these musical

performance evaluations as a means to justify the importance of music education in schools. Choral directors might believe standardized music performance evaluation validates their effectiveness as teachers.

The present study was pursued to explore what influences adjudicators' decisions regarding music performance quality. Identifying and acknowledging the influences on adjudicators' ratings when they assess a music performance may have positive implications for future adjudication training programs and for state-mandated music performance evaluation processes in general.

### **Purpose of the Study**

The purpose of the present study is to examine influences on choral music adjudicators' rating decisions related to the quality of a choral ensemble's performance. Music performance adjudication is a common experience for choral teachers and students; however, music educators have expressed concerns regarding adjudicator bias (Corbin, 1995; Fox, 1990; Robinson, 1990; Ward, 2013). The researcher intends to uncover factors that may influence decisions made by adjudicators when adjudicating choirs. The impetus for the present study stemmed from the researcher's desire to understand better choral music performance adjudicators' perspectives about what influences their rating decisions of choral music performance and to advocate for improved practices in performance evaluation. Implications from this study might influence teacher and program evaluation, professional development, policy makers' decisions, and ease chorus teachers' tensions or anxieties related to music performance evaluation (Batey, 2002). Results from the present study may provide stakeholders with

information about what might influence adjudicators' rating decisions of a choral music performance and help them to understand better the music performance adjudication process.

### **Research Questions**

Specific research question for the present study are:

1. What factors influence adjudicators' decisions when adjudicating choirs?
2. Do differences exist among adjudicators' influences on rating decisions by years of adjudication experience?
3. Do differences exist among adjudicators' influences on rating decisions by academic training?

## **CHAPTER II**

### **RELATED LITERATURE**

#### **Introduction**

The purpose of this literature review is to position the present research study into the larger body of research focused on the evaluation and assessment of choral musical performance. In this chapter, factors will be examined regarding music performance adjudication as described by McPherson and Thompson's (1998) Process Model of Assessing Musical Performance. Research in music performance assessment exists among many musical domains: from instrumental to vocal, solo-based to ensemble-based, audition-based performance to public performance, school music programs to professional music programs, and non-competitive festivals to competitions. The present research study will focus mostly in areas of research regarding vocal and choral ensemble assessment. Other musical domains provide rich information regarding music performance assessment, some of which will be included.

The first factor in assessing musical performance is performance context. This factor influences the purpose of the assessment and the musical performance, which subsequently influences the methods used for music performance evaluation. The second factor is musical and nonmusical factors. These are inherent in music performance and play an important role in how musical performance is perceived and evaluated by adjudicators.

The third and fourth factors are characteristics of the performer and evaluator, respectively. Characteristics of the performer logically contribute to the musical performance, as attributes of the performer provide aural and visual stimuli for the observer/listener. Similarly, characteristics of the evaluator may contribute to how an adjudicator assigns a rating to a musical performance. Factors within 'evaluator characteristics' are the primary areas of focus for the present research study. More specifically, the present research study seeks to investigate evaluator characteristics in the context of choral ensemble performance assessment.

The fifth factor is the selection of evaluative criteria, the evaluation instrument, and subsequently the evaluation process. A great deal of research is available in this area. In general, researchers and practitioners have worked to determine important aspects related to musical performance that resulted in the selection of evaluative criteria. These criteria were used to create evaluation instruments in the form of performance rubrics and ratings forms. These instruments provided a structure for the evaluation process and aided adjudicators in their determination of performance quality. The present research study will explore the available research regarding the selection of evaluative criteria and instruments for music performance adjudication.

All of the factors mentioned above influence the central factor of any music performance assessment, the musical performance itself, "the most critical determinant of any assessment" (McPherson & Thompson, 1998, p. 14). The present researcher intends to review available research regarding the assessment of musical performance and address possible concerns regarding evaluator characteristics.

### **A Brief History of Music Festivals and Competitions**

Rohrer (2002) provided a review of the history of music competitions and festivals in the United States. Music Supervisor's National Conference (MSNC), the forerunner of today's National Association for Music Education (NAfME), established a committee to oversee the National Band Competition in Chicago in 1923, which was the first national band contest in the United States. Shortly after, in 1924, the MSNC committee chose to standardize the competition with hopes to influence smaller and more localized music competitions across the country.

Music Supervisor's National Conference (MSNC) formed the National School Band Association (NSBA) to govern band contests. In 1926 MSNC held its first general session on music contests. One major point of consensus of the 1926 general session on music contests was that music contests stimulated teaching excellence in the schools.

The first National School Orchestra Contests occurred in 1929. Rohrer (2002) discussed the following:

By 1931, colleges and state teachers' associations sponsored contests all over the nation, and the national contest, although it remained in existence, moved toward state, regional, and local sponsorship . . . In 1937, the national sponsorship of music contests was officially replaced by sponsorship from ten geographic regions. Band, orchestra, and choral associations became responsible for their own affairs. (p. 39)

The Music Educators National Conference (MENC) reorganized national music competitions and festivals in 1948 followed by the National Interscholastic Music Activities Commission in 1952. Following the National Interscholastic Music Activities

Commission in 1968, the competition format returned to localized control and varied from state to state.

In a 2002 review of research, Rohrer revealed that music competition had been consistently a controversial topic among music educators, administrators, and researchers. In addition to describing the evolution of music competitions and festivals, researchers' discussions of the pros and cons of such practices were detailed. Rohrer cited Neil's (1945) study to provide support for music competitions, "directors valued contests for (1) the use of better music, (2) the improvement of instrumentation, (3) increased interest in school music by parents and students, (4) adjudicators' comments, and (5) the opportunity for students to hear other groups" (p. 41). Conversely, Rohrer cited many studies that recommended not to have competition in music. For example, Rohrer cited Coleman (1976), who compared music competition to athletic competition,

The competitive situation is one in which reinforcement is prescribed on the basis of a subject's behavior relative to that of other individuals; while the cooperative or less-competitive situation involves working in harmony to achieve a mutually agreeable end. The person engaged in competition is concerned with winning, while the goal of winning need not be present under cooperative conditions. (p. 41)

Rohrer commented on another area of concern regarding music competition, "Despite a rating system that replaced a ranking system, success at competition-festivals became a unit of measure for many communities and school systems for evaluating music programs and directors" (p. 42). Rohrer concluded that the individual [music director] should decide whether or not he or she would choose to participate in music competitions.



Payne (1997) stated, “Although a paucity of historical research exists, researchers generally refer to the National School Band Contest of 1923 as the official beginning of the competition movement in the United States” (p. 1). Payne continued to provide a detailed description of what researchers argued were the purposes for music competition-festivals. One theory was that music competition-festivals were influenced by immigration. Another theory was that music competition-festivals were influenced by the music industry to increase sales. Other theories suggested that military bands influenced competition-festivals.

Eventually criticism concerning the contest format, adjudication process, and focus on winning the competition lead to a shift in philosophical motivation (Payne, 1997). The names of these events changed from ‘contests’ to ‘competition-festivals;’ Rankings (i.e., first place, second place, third place, and so on) were replaced with performance ratings (i.e., Superior, Excellent, Good, Average, Poor). Sight-reading performance was added to reinforce music teachers’ instruction of music reading in school music programs. Adjudicator regulations were established including a requirement to have three performance adjudicators and one sight-reading performance adjudicator—a total of four adjudicators.

Many state music educator associations now host some form of music performance contest or festival for school music programs. Music contests and festivals occur typically in the spring semester of an academic school year. Traditionally, state music educator associations named their annual music events contests, festivals, or competitions (Payne, 1997). Payne stated, “the terms ‘contest,’ ‘festival,’ and

‘competition-festival’ remain relatively interchangeable in professional literature” (p. 3). During the 2008-2009 school year, the North Carolina Music Educators Association (NCMEA) renamed music festivals to be Music Performance Adjudications (North Carolina Music Educators Association, 2009). Other state music associations also adopted the term Music Performance Adjudication while others chose to use Music Performance Assessment (see Alabama Vocal Association, 2015; Florida School Music Association, 2015).

Each of these terms (i.e., state music festival, music performance adjudication, and music performance assessment) share similar expectations—a music ensemble performs for a group of judges who assigns a rating and provides constructive comments. Bergee and Platt (2003) provided a description of the adjudication process, which included a rubric, performance criteria, and comments/suggestions for the performing ensemble:

Judges used a rubric on the form to indicate level of achievement on various performance criteria (e.g., tone quality, intonation, technique, phrasing). They also wrote comments and suggested strategies for improvement under the broad categories of the fundamental technique and musical effect. (p. 346)

As music competition-festivals grew in popularity and in number, concerns over the way adjudicators evaluated performance became evident. Researchers have noted problems with the consistency of adjudication (Bergee & Platt, 2003) and specifically with the reliability of adjudicators’ scores among one another and across events.

### **Process Model of Assessing Musical Performance**

In the present literature review, the assessment of Music Performance Adjudication (MPA) utilizing the Process Model of Assessing Musical Performance (PMAMP) developed by McPherson and Thompson (1998) will be described. McPherson and Thompson (1998) adapted Landy and Farr's (1980) Process Model of Performance Rating (PMPR) to create the PMAMP.

Landy and Farr (1980), in their seminal work, developed the Process Model of Performance Rating (PMPR) in response to researchers' suggestions, "a model of some sort is necessary before any significant advances can be made in understanding judgmental performance measures" (p. 94). The PMPR consisted of three components [factors]: (a) position characteristics, (b) organization characteristics, and (c) the purpose for rating. The 'purpose for rating' component [factor] included a subcomponent, the rating process. The 'rating process' included two subsystems: (a) the cognitive process of the rater [adjudicator], and (b) the administrative rating process of the organization.

Music performance adjudication is a highly complex system comprised of interactions among many factors. Landy and Farr (1980) characterized music performance rating events, ". . . we are dealing with a constellation of activities rather than with single physical or mental operations in isolation" (p. 74). Many interacting aspects make the evaluation of the MPA process challenging. Music performance adjudication, at a basic level, might include these variables:

- (a) the roles (rater and ratee), (b) the vehicle (the rating instrument), (c) the rating context (the type of organization, the purpose for rating, etc.), (d) the rating process (administrative constraints, individual rater strategies, etc.), and (e) the

results of the rating (raw and transformed performance information, actions based on that information, etc.). (Landy & Farr, 1980, p. 73)

McPherson and Thompson's (1998) PMAMP was based on Landy and Farr's (1980) PMPR by utilizing its structure and salient features. Both models are similar in that they were designed to guide researchers toward novel research studies to investigate the potential causes of [adjudicator] bias that might impact the assessment of music performance in formal settings. McPherson and Thompson's (1998) PMAMP was an extension of Landy and Farr's (1980) PMPR by including research that occurred after 1980 and before 1998, and by adding components [factors] that further described the assessment of music performance.

McPherson and Thompson's (1998) PMAMP was different from Landy and Farr's (1980) PMPR. In addition to the components [factors] previously mentioned (i.e., position characteristics, organization characteristics, and the purpose for rating) the PMAMP included (a) the performance context, (b) musical factors, (c) non-musical factors, and (d) the musical performance, as components [factors] that impact the assessment of musical performance.

The PMAMP was different from the PMPR in regards to the terms used to describe each component [factor]. Landy and Farr (1980) used the term 'component' to represent each aspect of their model. McPherson and Thompson (1998) used the term 'factor' to represent each aspect of their model. Landy and Farr used the terms rating, rater, and ratee, to characterize the process, performer, and evaluator, respectively.

McPherson and Thompson used the terms evaluation process, performer, and evaluator, among other terms that seemed more specific to the present researcher.

McPherson and Thompson's (1998) PMAMP included the salient features of Landy and Farr's (1980) PMPR, however; the PMAMP was (a) more comprehensive with the addition of new factors and (b) seemed more precise by researchers' use of specific terminology. McPherson and Thompson's (1998) PMAMP is the most current and comprehensive model for evaluating the MPA process (a determination made by the present researcher). The present research study will utilize this model as a guide to explore available research regarding the MPA process. See Figure 1.

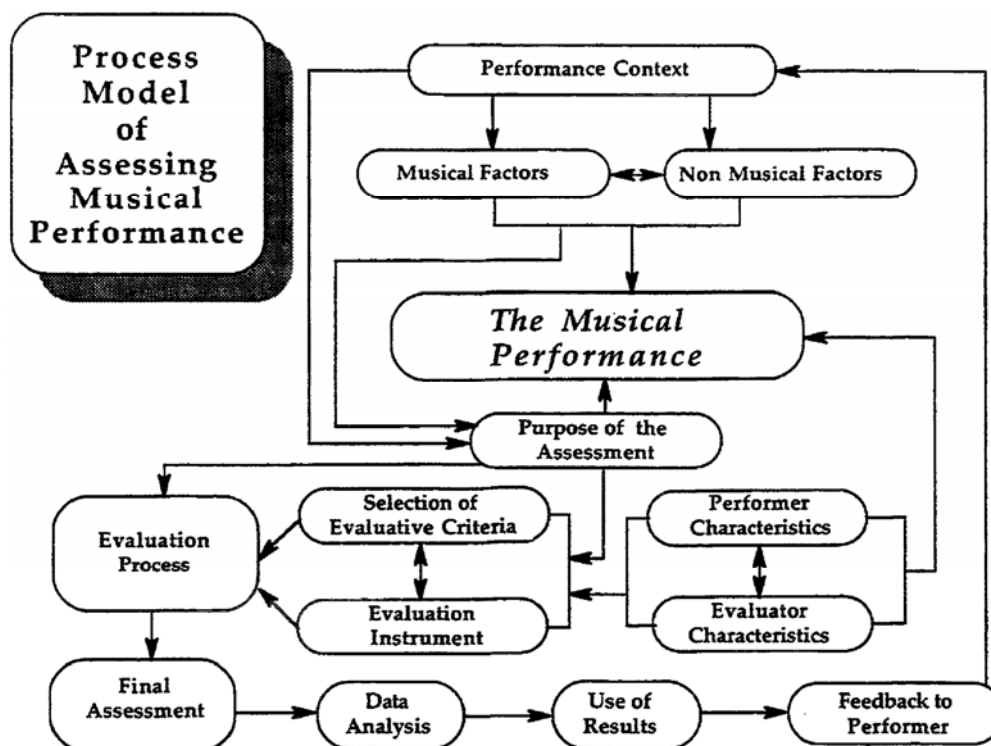


Figure 1. Process Model of Assessing Musical Performance. From "Assessing Music Performance: Issues and Influences," by G. E. McPherson and W. F. Thompson, 1998, *Research Studies in Music Education*, 10, p. 13. Copyright 1998 by SAGE Publications. Reprinted with permission.

### Performance Context and Purpose of the Assessment

Performance context is a necessary consideration when evaluating a music performance. Many factors play important roles: (a) the purpose of the event, (b) the type of performance, (c) performance proportions, (d) and the performance environment (McPherson & Thompson, 1998). Included within these factors are sub factors. For example, within the purpose of the event one must consider if the performance is part of a competition, festival, recital, audition, and so on. McPherson and Thompson (1998) stated:

First, the *purpose of the assessment*—whether the musician is performing in a music competition, festival, end of semester recital/examination, audition for an ensemble, or even participating in a music research project—strongly influences the way a judge will listen to, and therefore evaluate, a musical performance. (p. 12)

For choral directors and singers to have positive educational experiences at MPA events, coordinators might need to inform adjudicators as to the purpose of the event. Ekholm (2000) and Ford (1999) found that when adjudicators were asked to rate musical performances, there were significant differences in the results based on adjudicators' backgrounds and areas of expertise. Both studies provided evidence that adjudicators' understanding of performance context and the purpose of assessment influenced their ratings.

Ekholm (2000) investigated the effect of singing modes and seating arrangement on choral blend and overall choral sound. In this study, choral conductors ( $n = 37$ ), voice instructors ( $n = 33$ ), and instrumental (non-vocal) musicians ( $n = 32$ ) rated choral

performances in four experimental conditions: “(1) soloistic singing/random seating; (2) blended singing/random seating; (3) soloistic singing/acoustic seating; and (4) blended singing/acoustic seating” (p. 126). Results from the study indicated that voice instructors favored soloistic singing over blended singing and conversely choral conductors favored blended singing over soloistic singing. Additionally, both voice instructors and choral conductors favored the acoustic seating condition over the random seating condition. None of the participant groups (i.e., voice instructor, choral conductor, instrumental musician) were provided information regarding how they were to listen to and evaluate each choral performance, aside from the music performance evaluation form derived from previous research (Cooksey, 1977). Each participant rated each performance based on the evaluation form and his or her respective preferences. Results from Ekholm’s study suggested that adjudicators of different backgrounds and areas of expertise arrive at performance assessment differently. Although adjudicators in Ekholm’s study were provided the same adjudication form, they were not informed of the purpose of the assessment and therefore assigned a rating based on their assumptions and areas of expertise.

Ford (1999) investigated participants’ preferences regarding choral tone quality. Participants included undergraduate music majors with choral or vocal training ( $n = 47$ ), undergraduate music majors with instrumental training ( $n = 47$ ), and undergraduate students without any music training ( $n = 43$ ). The researcher recruited eight graduate voice students to perform and record four excerpts from four choral music selections. Each singer recorded each musical selection twice: (a) one with a full soloistic tone

quality and (b) another that was more blended (a unified ensemble sound). The more soloistic tone resulted in a stronger upper resonance in the singer's formant range. The blended (unified) tone reduced the singer's formant resonance. Each participant was randomly placed into one of two listening groups. Participants were asked to note which recording excerpt they liked best. Ford indicated that there was a significant difference ( $p < .000$ ) between the less-resonant and more-resonant responses. Participants preferred the blended (more unified) and less-resonant choral tone to the soloistic and more-resonant choral tone regardless of their major and prior musical training. Ford's analysis of participants' responses indicated that there was a significant difference ( $p < .000$ ) between the choral/vocal group and the instrumental group and a significant difference ( $p < .001$ ) between the choral/vocal group and the no-music-training group. Similar to Ekholm's (2000) study, adjudicators in Ford's study were provided a rating form and instructed to evaluate each musical performance based on their preferences. Ford (1999) found that when adjudicators are asked to choose based on preference, there are significant differences between adjudicators with various backgrounds and areas of expertise.

Taking into consideration choral directors and singers' expectations of MPA is important. Choral directors' reasons for participation in competitions and festivals as well as singers' motivations for participating have been investigated. Millard (2014) examined reasons choral directors take their ensembles to competitions. One of Millard's research questions was, "What are the reasons high school choral educators take their ensembles to competitions?" (p. 67). The three most popular reasons selected were "to



increase student work ethic/motivation” ( $n = 140$  or 79%), “to hear the judges’ feedback” ( $n = 123$  or 69%), and “to increase attention to musicianship” ( $n = 107$  or 60%) (p. 67-68). Millard determined that music educators participated in competition to motivate students toward musical achievement goals. The researcher also determined that music educators perceived that their students preferred to attend competitions to attain ratings and recognition. Millard’s study provided insight into what might motivate choral directors and singers to participate in MPA.

Ekholm’s (2000) study revealed that music adjudicators with solo-voice instruction expertise preferred the soloistic choral tone while choral conductors preferred a blended choral tone. Ford’s (1999) study suggested that undergraduate music and non-music students preferred a more blended choral tone to the soloistic choral tone. Results from both studies indicate that adjudicators’ ratings of musical performance might be influenced by a variety of different factors.

### **Type of Performance**

Adjudicators’ prior knowledge and assumptions regarding a performance medium might influence how they perceive and subsequently rate a musical performance. Performance types might include performing from memory (to reproduce a selection of music learned from notated music), sight-reading or sight-singing, performing by ear (to reproduce aurally a passage of an existing song heard repeatedly), improvisation, and so on. Performance types might also include performance on various instruments or voicing. Bias effects (regarding performance medium) among adjudicators might also

play a role in a musical evaluation, “. . . the *type of performance* that is being assessed will affect judgments” (McPherson & Thompson, 1998, p. 12).

Cassidy and Sims’s (1991) study provided evidence that adjudicators’ bias and prior knowledge regarding performers influenced performance ratings. Adjudicators’ expectations might be influenced by prior knowledge of the performer in the form of a stigmatized label. “The biasing effect of labels is clear in the results of several studies” (Cassidy & Sims, 1991, p. 24). Radocy (1976) tested the effect of adjudicator bias and determined that adjudicators’ perception of performance might be influenced by authority figures’ opinions. At the conclusion of his study, Radocy stated, “Bias effects with differing performance media should be explored” (p. 128). McPherson’s (1995) work validated the presences of multiple types of performance and evaluation and suggested different methods of assessment for each type of performance.

Cassidy and Sims (1991) investigated the effects of special education labels on music performance evaluations of a handicapped youth choir. In their study, participants rated a musical performance of a youth choir whose members were disabled. Participants were divided into four groups: (a) labels/visual, (b) labels/no visual, (c) no labels/visual, and (d) no labels/no visual. Participants in the first group received a label prior to viewing the video performance, which informed them that the choir’s membership consisted of young people with varying physical and cognitive disabilities. Participants in the second group were also informed of ensemble members’ disabilities, however; they only listened to the performance. Participants in the third group were not informed of choral members’ disabilities and watched the video of the performance. Remaining

participants listened to a recording of the performance without being informed of choir members' physical and cognitive conditions. "Results indicated a significant three-way interaction among label condition, presentation mode, and age" (p. 23). Mean performance ratings of participants who did not receive the label or video information rated the musical performance lower than mean performance ratings of participants who did receive the label or viewed the performance. Results from Cassidy and Sims's study suggested that adjudicators' perceptions of music performance quality might be influenced by information regarding the performers that was learned prior to viewing or listening to the performance. Results also indicated that adjudicators were influenced by the visual stimuli via the video condition suggesting possible bias.

Radocy (1976) investigated the effect of authority bias on participants' judgment of musical events. In his study, undergraduate music students evaluated identical musical performances under different bias conditions. Music authorities structured biased conditions and provided bogus information regarding the performers and composers. Participants tended to articulate their judgments in agreement with the bogus-bias produced by the authorities. Results from Radocy's work suggested that people might be influenced by the view or opinion of a perceived authority in their domain. Radocy intentionally manipulated participants' perceptions of musical performance by introducing false and inaccurate information via musical authorities.

McPherson (1995) examined five measures of musical performance. Participants included 101 high school clarinet and trumpet students who were preparing for the Australian Music Examinations Board (AMEB) Grades three to six final examinations.

Participants were divided into two groups based on age and ability: (a) Group One participants ranged from 12 to 15 years of age undertaking the ‘upper beginning’ level AMEB exam and (b) Group Two participants ranged from 15 to 18 years of age undertaking the ‘lower developing’ level AMEB exam. Prior to the main study, participants completed “the measures of the ability to sight-read, play from memory, play by ear and improvise, along with a specially designed questionnaire . . . used to obtain information on 16 variables believed to influence the development of these skills” (p. 145). The measures of ability consisted of three tests: (a) Test of Ability to Play From Memory (TAPFM), (b) Test of Ability to Play by Ear (TAPE), and (c) Test of Ability to Improvise (TAI) (McPherson, 1993). Scores from each test were compared with scores obtained using the Watkins-Farnum Performance Scale for sight-reading ability and AMEB awards for rehearsed musical performance. Finally, McPherson examined 16 variables regarding participants’ musical background to determine the degree of correlation with the five measures of musical performance. McPherson concluded the results of the study provided evidence of the validity of the three new music performance measures. He also concluded that the magnitude of correlation (between participants’ background and each music test) was greater for Group Two participants than Group One.

### **Musical Factors and Non-Musical Factors**

#### **Performance Environment**

Performance environment may influence music evaluation. McPherson and Thompson (1998) stated, “Finally, both performance and assessment may be influenced

by the *performance environment*, such as the size and acoustics of the performance space and the equipment available for the performer” (p. 14).

In general, the performance environment affects music performance adjudicators. There are a wide variety of contextual conditions among choral performance adjudication situations. There are also inevitable differences among adjudicators as they might vary in age, health, experience, and personal needs. Additionally, adjudicators’ music listening experience and their ability to concentrate their attention are affected by many variables including but not limited to fluctuations in mood, time of day, illness, fatigue, room conditions, visual aspects, personal characteristic, personal needs, how the choir looks, and how the choir carries themselves (Bergee, 2007; Bergee & Platt, 2003; Bergee & Westfall, 2005; James et al., 1984; McPherson & Thompson, 1998).

Music performance adjudication events vary among districts, regions, and states. Some events might last half a day while others might be three to four days in length, or longer. A typical day of adjudication might begin at eight o’clock in the morning and last until six o’clock in the evening, with a different choir performing every fifteen minutes. Depending on the local school system, non-music-related statewide testing, or the myriad of school-based activities in which the school-age music students might participate, scheduling conflicts may cause some music performance adjudication events to start early in the day or last through late in the evening. Some events take place in churches while others take place in school auditoriums. Adjudicators might sit on padded chairs, hard wooden or plastic chairs, or stools. Adjudicators might also be oriented in the listening space at ground level, on an elevated platform, or in the balcony of an

auditorium. The adjudication season might begin in late February and end in mid-April when the weather shifts from winter, with possible snowy or icy conditions, to spring, with seasonal allergens and strong thunderstorms.

### **Performance Proportions**

Performance proportions might include a solo-based or ensemble-based performance. This indicates the number of performers being observed in the musical performance. It is reasonable to assume that performance proportions of the ensemble affect visual and aural stimuli—the size of an ensemble will provide a unique aesthetic. These proportions might affect adjudicators' assessments. “Third, *performance proportions* (whether the performance involves a soloist or ensemble) affect assessment both by shaping the aesthetic goals of a performance, and by constraining extramusical influences on assessment” (McPherson & Thompson, 1998, p. 14).

Visual aspects, including size of ensemble played a significant role in how adjudicators rated choirs at music performance adjudications (Bergee & Platt, 2003; Bergee & McWhirter, 2005; Bergee & Westfall, 2005). Bergee and McWhirter, (2005) analyzed data from a solo and ensemble music festival ( $N = 3,853$ ) and found significant differences among the main effects of type of performance (solo and ensemble) and performance medium (vocal and instrumental). The study was an extension of Bergee and Platt's (2003) study that analyzed 7,355 solo and ensemble vocal and instrumental outcomes from two state music performance adjudication events. The assistant executive director of the sponsoring organization provided the data (music performance ratings) used for the study. Bergee and Platt coded the data: (a) time of day in eight categories,

(b) type of event in two categories, (c) performing medium in to two categories, and (d) school size in five categories. The researchers calculated separate analyses of variance (ANOVA) for each music performance adjudication event with “time of day (eight levels), type of event (solo or ensemble), performing medium (vocal or instrumental), and school size classification (five levels) serving as independent variables, and overall rating received as the dependent variable” (p. 346). Bergee and Platt (2003) found, “statistically significant differences in the main effect of time of day [ $F(7, 7196) = 3.29, p < .001$ , partial  $\eta^2 = .003$ ], type of event [ $F(1, 7196) = 10.68, p < .001$ , partial  $\eta^2 = .001$ ], and school size [ $F(1, 7196) = 10.68, p < .001$ , partial  $\eta^2 = .009$ ]” (p. 342).

Non-musical factors influence how audiences perceive aural information.

McPherson and Thompson (1998) provided a rich review of research concerning issues and influences regarding the assessment of musical performance. One major component was non-musical factors’ influence on adjudicators’ evaluations of performance quality, “A variety of nonmusical factors can have a profound influence on the assessment” (p. 14).

### **Characteristics of the Performer and the Evaluator**

#### **Appearance**

How might non-musical visual information influence the way in which adjudicators rate a musical performance? Researchers in the domains of psychological and cognitive sciences, and music perception and cognition, provided evidence that peoples’ perceptions of musical performance are strongly influenced by a combination of what they see, their expectations, a variety of non-musical factors, and auditory

information. Mitchell and MacDonald (2016) found that visual information plays a critical role in one's assessment of a musical performance. "While listeners maintain they are focused on sound when evaluating performers, recent music research suggests that audiences use visual information to complement their audio experience and that visual information alone can transmit musical intentions and quality" (p. 1). It is interesting to know that what one sees does in fact influence what one hears.

Tsay (2013) investigated sight-over-sound in the judgment of a musical performance. In this study the researcher conducted seven experiments that challenged the notion that expert music judges attend to auditory information more than visual information. Tsay found that "people actually depend primarily on visual information when making judgments about music performance" (p. 14580). Tsay provided evidence that visual stimuli might affect adjudicators' ratings of musical performance.

In experiment one, Tsay (2013) explored participants' core beliefs about music. Participants were given the opportunity to win a cash prize if they could guess the winner of a live music competition. Participants had the choice of an audio-only recording, video-only recording, or an audio-video recording. Participants chose sound recordings (58.5%) significantly more than those who chose video recordings (14.2%). Other participants (27.4%) chose the audio-visual recordings, which was a significant difference from the video-only recordings. From experiment one Tsay concluded "People have the intuition that sound is a more revealing channel of information in the domain of music and the recordings with both visual and auditory output offer additional and more relevant information that better approximates the conditions..." (p. 14581).



In experiments two through five, participants reported that sound mattered more in their evaluations of musical performance. Although participants reported they only attend to sound, results from each experiment indicated that they relied more on visual stimuli when making their judgment. Tsay (2013) reported,

In experiment 2 . . . Although 83.3% of participants reported that the sound mattered most for their evaluations of music performance, these same participants were significantly more likely to identify the winners when they were presented with only visual components of the performances,  $t_1(105) = 12.07, P < 0.001$ ; Cohen's  $d = 1.18$ . (p. 14581)

Experiment three utilized three conditions: (a) video-only, (b) sound-only, (c) video-plus-sound. Participants in the video-only condition were able to identify winners of the music competition more often (46.4%) than those in the other two conditions. In experiment four, Tsay reported, "Despite musicians' training to use and value sound in their evaluations, only 20.5% of experts identified the winners when they heard sound-only versions of recordings" (p. 14582). In experiment five, participants performed significantly better with video-only stimuli than sound-only stimuli ( $t_1(61) = 4.48, P < 0.001$ ; Cohen's  $d=1.20$ ). Experiments six and seven focused on performers' movement and gesture, which are visual elements of performance. Results from experiment six indicated that participants scored significantly better than chance (48.8%) at identifying the winners of a music competition. Experiment seven focused on participants' ability to identify the "most confident, creative, involved motivated, passionate, and unique performer" (p. 14583). Participants ( $n = 262$ ) viewed video-only and sound-only stimuli of three contest finalists. Results from this experiment indicated that "Creativity,

involvement, motivation, passion, and uniqueness were significantly more salient through visual cues rather than through sound” (p. 14583).

In all the experiments, Tsay (2013) revealed that people intuitively knew that sound was imperative for interpreting information in music; however, visual aspects of musical performances affected judges’ responses. “These results demonstrate how visual information, the information generally deemed as peripheral in the domain of music, can be outweighed when such inclination is neither valued nor recognized” (p. 14582).

### **Conductor’s Visual Appearance**

Morrison, Price, Geiger, and Cornacchio (2009) discovered that audience members perceived ensembles to be more aurally expressive when their conductor appeared more visually expressive. In their experiment, Morrison et al. controlled for the sound of the choir by using the same audio track for both the ‘less expressive’ and ‘more expressive’ conducting conditions—the audio did not change. This study provided evidence that visual information, in the form of a visually expressive conductor, informed listeners’ perception of an ensemble’s music performance quality.

Other researchers explored the line of study regarding audiences’ perception of choral ensemble quality and conductor’s visual performance. Madsen (2009) sought to investigate the effect of presentation mode (i.e., visual and aural) on evaluations of conducting and choral ensemble performance. The researcher wanted to know if visual aspects of conducting increased or decreased audience members’ ratings of an ensemble’s performance. Results from the study indicated that different styles of

conducting behavior might influence audience perceptions—poor conducting might influence ratings more so than good conducting.

Madsen (2009) found similar results as did Morrison et al. (2009); the more visually expressive the conductor appeared the higher the ratings for the ensemble's performance. The most important discovery, however, was that evaluators were more sensitive to the 'poor conducting' condition than to the 'good conducting' condition. A notable factor regarding participants in Madsen's study, they were music professionals with advanced degrees in conducting and music education. This fact suggests that there might be bias among music professionals regarding the appearance of an ensemble's conductor including his or her conducting style.

Napoles (2013) found that the appearance of the conductor influenced audience members' perceptions of performance quality. Napoles also discovered that presentation mode, the way in which the choir is presented on stage, affected audience members' perception of performance quality. Napoles's study supports the notion that the more expressive a conductor looks while conducting, the more expressive the choir might seem during performance.

### **Ensemble Movement**

According to Juchniewicz (2008), ensemble members' movement during an adjudicated performance significantly increased audiences' ratings of the performance. More specifically, as physical movement increased, so did ratings of phrasing, dynamics, sense of rubato, and overall musical performance. Furthermore, Juchniewicz found no

significant differences between levels of training (raters' training) or gender (male or female). This information suggests that the choir's movement influenced the evaluators.

Researchers performed studies at choral competitions regarding ensemble movement. Researchers suggested that ensemble size dictates the appropriate amount of movement that an ensemble might use to create an aesthetic effect and increase quality ratings from adjudicators. McPherson and Thompson (1998) cited Morgan and Burrows (1981), "For choral competitions, research suggests that large groups should restrict the degree of physical movement to avoid appearing too 'busy' on stage, whereas smaller groups can afford a larger range of movements during a performance" (p. 14). The visual influence of movement affected adjudicators' perception of overall performance quality.

### **Race Perceptions**

Race, as defined by the Merriam-Webster (1983) dictionary, is "any of the major biological divisions of mankind, distinguished by color and texture of hair, color of skin and eyes, stature, bodily proportions, etc." (p. 1484). Outward appearance is one of the simplest observations anybody can make in any situation or context. Researchers found that racial stereotypes and perceptions significantly affect judgments regarding musical performance quality (Cheek, 2007; Elliot, 1995/1996). Researchers provided evidence that peoples' perceptions of race inform their expectations regarding musical performance (Brown & Novak, 2007). This information supports the notion that adjudicators might be bias towards specific groups of people and this bias might influence ratings of music performance quality.

Elliot (1995/1996) performed a study to investigate the possible effects of race and gender on evaluations of musical performance. He found that race played a significant role in adjudicators' evaluations of musical performance. "The main effect of race was found to be significant, with blacks scoring significantly lower than whites" (p. 53). He also found that gender stereotyping played a role in adjudicators' evaluations of musical performance. He discovered an interaction effect between gender and race, "For the black performers, males tended to score the lowest; whereas among whites, the females tended to score lower" (p. 53).

Cheek (2007) investigated adjudicators' race and racial perceptions on choral ensemble performance. In this study, Cheek presented pictures of three choral ensembles; (a) an all-black choir, (b) an all-white choir, and (c) a mixed-race choir consisting of black and white singers. Participants in the study were provided a still image of one of the choirs (black, white, or mixed) and listened to a recording of a choir that may or may not have been representative of the images. Participants had to decide if the choir they heard was all black, all white, black and white, or other. Results from the study informed us that adjudicators were affected by their race and racial bias. In other words, the visual stimulus of the photo, which highlighted the possible racial makeup of the ensemble, influenced participants' perception of the audio information received from the audio recording.

### **Stage Presence and Personality**

First impressions matter in a musical performance. How a musical artist or ensemble presents themselves to an audience impacts their audiences' perception of the

entire performance. Platz and Kopiez (2013) found support for the idea that audience members' first impression of the performer or group has an influence on audience members' perception of musical performance. Many music teachers invest time on a 'stage entrance' routine for their students. Professional musicians and ensembles practice a routine for entering and exiting the stage. Certain music traditions expect particular entrance and exit behaviors before and following all musical performances. This non-musical behavior impacts how an audience, or adjudicator, will perceive the aural information presented.

Howard (2012) suggested that performance quality ratings are affected significantly by performance attire and stage deportment. Howard found that what a person wears when performing a musical selection influences adjudicators' ratings regarding the quality of the performance. Additionally, a performers' deportment, the manner of conducting or bearing oneself while on stage, influences adjudicators' ratings regarding the quality of the performance. In her study, participants assigned significantly higher ratings to audio-only evaluations. Performers' choice of attire and how they conducted themselves on stage influenced participants' evaluations during audiovisual conditions. These two non-musical visual aspects, clothing and conduct, influence how people perceived aural information.

### **Evaluators' Personality**

Does a person's personality contribute to how they listen to music? Thomas et al. (2015) study suggested a positive correlation between the personality trait 'Openness to Experience' and one's auditory discrimination ability. This finding suggested that a

person with higher levels of Openness possesses the ability to differentiate small and significant units of sound, a skill necessary for the assessment of musical performance. Might this personality trait affect an adjudicator physically and emotionally, resulting in a pleasurable experience? Colver and El-Alayli (2016) found that the personality trait Openness correlated positively with frisson—a term used to describe the emotional and physical nature of ‘shivers’ or ‘chills’ associated with a pleasurable musical listening experience. Logically, the personality trait Openness and frisson experience might influence adjudicators’ ratings of musical performance.

Thomas et al. (2015) conducted a study to explore peoples’ personality traits as predictors of auditory discrimination ability. A total of 184 college students at a large institution in the southeastern United States participated in the study. Most of the participants ( $n = 174$ ) chose to participate as part of a research option for a college course. Other participants ( $n = 10$ ) included music majors in undergraduate and graduate level degree programs in music. Each participant completed The Musical Ear Test (MET), the HEXACO model of personality, and a musical training assessment. The MET was designed to measure a person’s melodic and rhythmic discrimination ability. The HEXACO model of personality was designed to measure six factors of a person’s personality: (a) Honesty-Humility, (b) Emotionality, (c) eXtraversion, (d) Agreeableness, (e) Conscientiousness, and (f) Openness to Experience. The musical training assessment measured participants’ training and education in music. Results from the study indicated that a person’s personality affects his or her auditory discrimination ability. The personality trait, Openness to Experience, significantly predicted peoples’ auditory ability

( $p < .001$ ). No other personality trait in the study had a significant effect. The researchers defined Openness to Experience as, “a broad trait associated with aesthetic and creative interests” (p. 1). In their discussion, Thomas et al. stated, “Our study found that people high in Openness to Experience had significantly higher auditory discrimination ability” (p. 7). This research supports the notion that a person’s personality will inform his or her evaluations of a musical performance.

Colver and El-Alayli (2016) tested the notion that the emotional nature of frisson (i.e., pleasurable aesthetic chills) is somehow related to personality traits (i.e., openness to experience) and that one might be able to predict if a person might experience frisson during the music listening experience. A total of 113 persons participated in the study. All participants were psychology or music undergraduate college students and received course credit for their involvement in the study. Sixteen participants’ data were deemed unusable by the researcher, which resulted in a final total of 97 participants. Participants completed a demographic information form and the NEO-PI-R (Costa & McCrae, 1992), a standardized measure of five personality domains including (a) Openness, (b) Conscientiousness, (c) Extroversion, (d) Agreeableness, and (e) Neuroticism. Next, participants listened to selected musical excerpts while connected to a Galvanic Skin Response (GSR) Biofeedback Machine. Following the listening session participants completed a Likert-type questionnaire regarding their familiarity with the musical selections. Results indicated that Openness correlated positively with frisson experiences,  $r(95) = .45, p < .001$ . “Openness correlated with participants’ responses to NEO-PI-R item 188 [aesthetic chills] at a significant positive level,  $r(95) = .69, p < .001$ .



Item 188 also correlated significantly with number of frisson episodes participants experienced,  $r(95) = .29, p = .004$ " (p. 9). The researchers determined that the Openness was the only dimension on the NEO-PI-R that had a statistically significant correlation with frisson response. In their discussion the researchers noted,

Using a combined self-reported/physiological measure of frisson, the present study found that higher Openness scores on the NEO-PI-R correlated with a greater number of actual experiences of frisson (aesthetic chills) during a live music listening session. This result both substantiates an Openness-frisson relationship and sheds greater light on its complex nature. (p. 11)

Thomas et al. (2015) concluded that people with high levels of the personality trait Openness to Experience possessed significantly higher auditory discrimination ability. Colver and El-Alayli (2016) concluded that the personality trait Openness correlated positively with frisson experiences—pleasurable aesthetic chills. Music performance adjudicators with a higher level of Openness might possess high auditory discrimination ability and experience frisson while listening to a live performance. Conversely, evidence suggests that adjudicators with lower levels of Openness might possess lower auditor discrimination ability and may not experience frisson while listening to live musical performance. Adjudicators' auditory discrimination ability and frisson experience may influence adjudicators' ratings of musical performance.

### **Evaluators' Music Experience and Ability**

Does a person's experience and ability with music contribute to how he or she might rate a musical performance? Research in this domain seemed to indicate that experience aids in internal consistency of performance ratings (Kinney, 2009), however;

experience was not a factor in selecting music competition winners (Tsay, 2013). Teaching level seemed to influence adjudicators' judgments regarding intonation and tone (Hewitt & Smith, 2004), however; adjudicators' academic level correlated inversely with performance ratings (Howard, 2012). Experience in music related coursework seemed to increase participants' music vocabulary and auditory discriminating ability during a music listening test (Hair, 1981). Musicians were able to detect rubato significantly more than non-musicians when evaluating a musical excerpt (Johnson, 1996), however; sophisticated musicians were unable to detect differences across five different audio recordings spliced together (Madsen, Geringer, & Wagner, 2007). Prior research suggested musicians' and non-musicians' experiences and ability influenced their evaluations of musical performance, but there seemed to be slight discrepancies. One salient feature that seemed to be consistent was that expertise influenced consistency (Hair, 1981; Hewitt & Smith, 2004; Kinney, 2009; Johnson, 1996).

Kinney (2009) examined effects of music experience and excerpt familiarity on the internal consistency of performance evaluations. Participants ( $n = 131$ ) in the study included non-music majors ( $n = 63$ ), music majors ( $n = 42$ ), and music experts ( $n = 26$ ). Results indicated that participants' experiences in music and familiarity of musical excerpts significantly affected internal consistency of performance evaluations and that these two factors interacted significantly. Kinney's findings suggest that adjudicators' musical experience and music expert familiarity influence the consistency of performance ratings.

Tsay's (2013) study, regarding sight over sound in the judgment of music performance, determined that people's experience with music might not influence one's ability to select the winner of a music competition. Following the completion of all seven experiments in his study Tsay stated,

Experts were not significantly different from novices in their judgments of musical performance. Novices and experts are similarly below chance with sound recordings and at chance with recordings with both video and sound. Novices and experts also paralleled each other in their use of different cues to arrive at the competition outcomes made by the original judges, with no significant differences though the sound-only recordings,  $t(95) = 0.85, p = \text{n.s.}$ ; the video-plus-sound recordings,  $t(106) = 1.68, p = \text{n.s.}$ ; nor the video-only recordings,  $t(81) = -0.12, p = \text{n.s.}$  (p. 14582)

Hewitt and Smith (2004) investigated the influence of teaching-career level and primary instrument on music teachers' assessment of musical performance. A total of 150 individuals participated in the study. Participants included in-service teachers ( $n = 33$ ) and undergraduate music education students ( $n = 117$ ). Undergraduate music education majors were divided into two groups: a lower-division group ( $n = 49$ ) and an upper-division group ( $n = 68$ ). Participants' principal instrument included brass ( $n = 48$ ), woodwind ( $n = 47$ ), percussion ( $n = 9$ ), strings ( $n = 30$ ), voice ( $n = 11$ ), and piano ( $n = 5$ ). The researchers employed a causal comparative design to examine relationships among independent variables (i.e., teaching level and primary performance instrument) and the dependent variable (i.e., scores on the Woodwind Brass Solo Evaluation Form). For independent variables, participants provided information regarding their teaching career and primary performance instrument. For the dependent variable participants rated six previously recorded musical performances using the Woodwind Brass Solo Evolution

Form. The researchers found a statistically significant two-way interaction effect between teaching level and performer [ $F(70, 22) = 1.47, p = .02, \eta^2 = 1.00$ ]. “Statistically significant main effects were found for the performer [ $F(35, 110) = 47.58, p = .00, \eta^2 = 1.00$ ] and teaching level [ $F(14, 278) = 2.54, p = .00, \eta^2 = .99$ ] conditions, but not for primary instrument [ $F(7, 138) = .33, p = .94, \eta^2 = .15$ ] (p. 321). Other significant differences existed among teaching levels (i.e., in-service, lower-division, upper-division) and music performance evaluation categories (i.e., intonation and tone). These findings suggest that teaching level influences a music evaluator’s judgments regarding music performance’s intonation and tone.

Howard (2012) investigated the effect of gender and academic level on music performance adjudicators’ evaluations of high school solo vocal performance. A total of 282 individuals participated in the study. Participants included high school chorus students ( $n = 153$ ), undergraduate music majors ( $n = 97$ ), and graduate level music majors ( $n = 32$ ). The researcher asked participants to evaluate high school solo vocalists’ overall musical performance quality using a six-point Likert-type scale (1 = *weak*; 6 = *strong*). Participants evaluated four solo vocalists (i.e., soprano, alto, tenor, bass) in four presentation conditions: (a) formal attire/formal department, (b) formal attire/casual department, (c) casual attire/casual department, and (d) casual attire/formal department. These performances were transformed into audiovisual and audio-only digital recordings then randomized to create five excerpts for participants’ evaluation. The researcher applied a three-way ANOVA “to examine adjudicators’ ratings of presentation conditions with formal versus casual attire as a function of adjudicator gender and academic level”

(p. 173). “A significant difference,  $F(2) = 7.838, p < .05, \eta^2 = .054$ , was identified for adjudicator academic level” (p. 173). “Results indicated a significant difference,  $F(2) = 5.081, p < .05, \eta^2 = .036$ , between the two types of stage department as a function of adjudicator academic level” (p. 175). High school level adjudicators assigned ratings higher than the undergraduate and graduate level adjudicators. There was an inverse relationship with academic level and assigned performance ratings—the higher the academic level the lower the performance rating.

Hair (1981) examined verbal identification of music concepts among children, adults, music and education majors, gender, and experience in music classes. The researcher was curious to know if experiences in music, among other factors, contributed to participants’ music vocabulary when discriminating differences during music listening. A total of 299 persons participated in the study, including children ( $n = 226$ ) and adults ( $n = 73$ ). Child participants’ school-grade-level ranged from first grade to fifth grade. They were divided into three ability groups: low, middle, and high. Fifty-six of the adult participants were elementary education majors enrolled in introductory music courses. The remaining 17 adults were senior music education majors. All participants performed a listening test that used the melody ‘Twinkle, Twinkle, Little Star.’ The melody was played a total of 10 times. Each presentation of the melody was different. Hair described each listening example,

The presentations of the melody were as follows: “normal” practice example, (1) loud, (2) soft, (3) fast, (4) slow, (5) high (one octave higher than original), (6) low (one octave lower than original), (7) sung (a cappella by the investigator), (8) harmonized, (9) parallel minor, and (10) with rhythmic change (dotted rhythm pattern throughout). (pp. 13–14)

Participants were instructed to write a word next to the number on the paper that best describes how the music sounded when compared to the ‘normal’ example. Analysis of test results showed significant differences between (a) children and adults, (b) music and education majors, (c) female adults and major, and (d) education majors at the beginning and end of music courses. Hair suggested the effect of music training influenced participants’ answers:

The effects of music training seemed evident on the last four items...In addition, education majors at the beginning of the course showed a great drop in scores for the last four items. Music and education majors at the end of their course, had consistently high scores on all items . . . (p. 20)

Results suggest that experience in music classes contributed to participants’ vocabulary use. Results might also suggest that experience in music contributed to participants’ ability to discriminate the differences between each playing of the melody.

Johnson (1996) investigated musicians’ and non-musicians’ ability to perceive and assess rubato in a musical performance. Essentially, the researcher sought to determine the difference between musicians and non-musicians regarding nuances of musical performance. A total of 96 persons participated in the study—musicians ( $n = 48$ ) and non-musicians ( $n = 48$ ). All musicians were undergraduate and graduate level university music majors. All non-musicians were undergraduate students enrolled in an elementary music education class for non-majors. Notable, was that non-musician participants claimed they could read a musical score, due to training in the non-major elementary music class. Participants were randomly placed into three condition groups: (a) Condition 1 received instruction and training on rubato detection, (b) Condition 2

received a short lecture regarding the definition of rubato, and (c) Condition 3 listened to Samuel Barber's *Air for Strings* and did not receive information regarding rubato. An expert panel listened to and assessed 13 professional recordings of Mozart's Concerto no. 2 in E-flat major for Horn and Orchestra. The highest and lowest rated performances were selected and used for the study, which resulted in four recordings. Participants were instructed to listen to, analyze, and evaluate each professional recording using the musical score, a Continuous Response Digital Interface (CRDI), and a Likert-type scale questionnaire.

Results indicated significant differences between musicians and non-musicians with regard to evaluations of rubato ( $F = 2.68, df = 3, 282, p \leq .05$ ), musicianship ( $F = 5.59, df = 3, 282, p \leq .01$ ), expression ( $F = 6.34, df = 3, 282, p \leq .01$ ), and tone quality ( $F = 12.90, df = 3, 282, p \leq .01$ ). (p. 92)

The researcher sought to determine significant differences that existed among the musician group. Johnson divided the musician group into two proficiency groups: (a) Group A—more musically proficient and (b) Group B—less musically proficient.

Results indicated significant differences between musicians in Group A and Group B with regards to evaluations of rubato ( $F = 7.87, df = 3, 138, p \leq .01$ ), musicianship ( $F = 3.05, df = 3, 138, p \leq .05$ ), expression ( $F = 3.98, df = 3, 138, p \leq .01$ ), and tone quality ( $F = 3.27, df = 3, 138, p \leq .05$ ). (p. 93)

Results from Johnson's study suggest that music performance adjudicators with more experience might be able to detect musical nuances with more sensitivity than less experienced musicians and non-musicians with music reading abilities.

Research supports the notion that music experience influences adjudicators' ratings of music performance. This research focuses on music experience as: (a)

instruction in music classes, and (b) teaching level. The present researcher recognizes an area of possible importance that is not represented in available research, adjudication experience. The present research study will investigate the area of years of adjudication experience.

### **The Musical Performance**

Geringer and Madsen (1998) conducted a study focused on musicians' attention when listening to music to determine whether musicians demonstrated consistent listening patterns across musical excerpts. A key feature of this study was that the researchers intentionally designed musical performances to be perceived as good and bad. A total of 48 music majors, undergraduate ( $n = 24$ ) and graduate ( $n = 24$ ), participated in the study. Musical excerpts came from the first and second phrases of Schubert's and Gounod's *Ave Maria*. Four soloists (i.e., soprano, tenor, violinist, and cellist) listened to and performed along with prerecorded accompaniment tracks. Each soloist attempted to perform his or her very best and worst with each recording. Good and bad soloist recordings were paired with and without accompaniment. Participants listened to and evaluated the final stimulus recordings. The researchers asked participants to rate performances using a ratings scale (phrasing/expression, intonation, rhythm, dynamics, tone quality) and give an overall rating. Results indicated that musician listeners consistently discriminated between the good and bad performances across all rating scales. Significant differences were found between the four timbres and the accompaniment conditions. Another result, "The presentation of performances with accompaniment was associated with higher ratings for three of the scales:



phrasing/expression, rhythm, and dynamics” (p. 531). The presence of accompaniment possibly contributed to higher ratings.

McPherson and Thompson (1998) postulated that in addition to a number of variables that might affect adjudicator’s evaluative decisions, “repeated hearings of the same piece may nurture a greater appreciation of that piece, which, in turn, might cast a more positive light on the performance itself” (p. 140). This characteristic, repeated listening, is prolific and unavoidable at music performance assessments due to a generally accepted requirement for choirs to sing musical selections from a state-approved list of pieces (North Carolina Music Education Association, 2015).

Adjudicators will most assuredly hear the same arrangements of pieces performed by multiple choirs over the course of a day, or longer.

### **Training**

Does training influence adjudicators’ ratings of music performance? Tsay (2013) suggested music training was not necessary in order to select music completion winners. Specific training in the form of adjudicator training influenced adjudicators’ ratings of choral performance (Ward, 2013). Experience was considered by some to be a form of training and has been shown to influence adjudicators’ expectations of music performance (Brown & Novak, 2007; Kinney, 2009; Mills, 1991). Researchers suggested that training in adjudication might increase inter-judge reliability (Mills, 1991; Ward, 2013). McPherson and Thompson (1998) called for more research to investigate adjudicators’ musical, social, and cultural biases. They suggested that all people in all places experience music listening differently. Peoples’ expectations about a

performance, prior to the listening experience, influence their evaluation of the performance (Duerksen, 1972).

Brown and Novak (2007) assessed the intrinsic impact of live musical performance. Using a paired questionnaire, one administered before the performance and another following the performance, the researchers investigated listeners' mental and emotional preparedness, and a range of reactions including: captivation, intellectual stimulation, emotional resonance, spiritual value, aesthetic growth, and social bonding. They explored areas of concern specific to how people perceive musical performance differently. Their research supported the argument that adjudicators' expectations of choral performances might influence evaluation. This research also supports the idea that adjudicator experience might act as a training mechanism for assessing music performance—building expectations for music performance.

Mills (1991) argued that years of adjudication experience provided enough training for end-of-year music performance assessment. Her study compared a holistic assessment method to a standardized method approved by authorities in test construction and assessment. The researcher used a mixed method design, which employed interviews and triangulation (Kelly, 1955) to determine discrete constructs used in holistic assessment methods. The researcher used statistical analyses to compare data from the two assessment methods. After constructs for the holistic method were determined, the researcher created a music assessment rubric to be used and compared with the approved standardized method. The researcher concluded, "Holistic assessment has advantages over segmented assessment" (p. 179). Furthermore, she argued that the list of constructs

found in her study could be used as a means for training adjudicators, possibly resulting in increased inter-judge reliability. Although she initially argued that experience was necessary enough for music assessment, she suggested that constructs produced from her study might be useful in future adjudicator training.

Ward (2013) conducted a preliminary study in North Carolina to investigate the development of error detection and assessment skills among choral performance adjudicators through adjudication training. The purpose of Ward's study was to see if choral performance adjudicators' error detection and assessment skills would improve after participating in a researcher-designed Choir Adjudicator Training (CAT) program. The researcher sought to: (a) examine and compare the inter-judge reliability of NCMEA adjudicators with the inter-judge reliability of undergraduate and graduate music majors and novice choral music educators, (b) examine the effectiveness of the CAT in the identification of errors in a choral performance, (c) identify perceptions of the CAT in preparing undergraduate and graduate music majors and novice choral music educators for the music education profession, and (d) identify perceptions of the large choral ensemble festival procedures of the NCMEA. One conclusion and subsequent recommendation from Ward's study was that adjudicators benefited from professional development and that the NCMEA High School Choral Section should provide such professional development.

Kinney (2009) examined the effect of music expertise and musical excerpt familiarity on internal consistency among adjudicators during music performance evaluation. Participants consisted of three groups: (a) non-music majors, (b) music

majors, and (c) experts. Participants in the non-music majors group had varied experiences with music ensembles while in high school—some were enrolled in a high school ensemble while others were not. Participants in the expert group consisted of graduate level music majors and music faculty. All participants listened to 45 short musical excerpts of three songs, completed an evaluation form for ‘accuracy’ and ‘expressiveness,’ and a demographic questionnaire. The musical selections consisted of three songs, two that were considered generally familiar (*Amazing Grace*, *America the Beautiful*), and one that was not (*Song Number Three*). The evaluation form was used to rate accuracy and expression via two Likert-type scales each with a one to seven scale with one representing low and seven representing high. “Results from this investigation suggest that internal consistency of performance evaluation is related to music experience and training” (p. 333). Additionally, musical excerpt familiarity influenced adjudicators’ internal consistency. Interestingly, music majors and experts rated familiar musical selections higher than unfamiliar selections consistently while non-music majors rated unfamiliar songs higher than familiar songs. The researcher suggested that academic instruction in music and music excerpt familiarity in the form of experience influenced music performance ratings.

Lehmann (1997) investigated “the relation between self-reported affective response to everyday life events and affective responses to music listening” (p. 84). The researcher examined the effect of participants’ formal music training on their responses to music listening. First, participants completed Larsen and Diener’s Affect Intensity Measure (1987, AIM). Lehmann explained:

Affect Intensity (AI) is defined as the general, stable characteristics of a person to experience various emotions in typical life situations more or less strongly, independent of frequency and objective severity of the emotion-inducing event and regardless of the hedonic tone or valence of the emotion. (p. 85)

After responding to the AIM, participants provided information regarding their prior musical training and demographic data. Then they completed a habitual and situational listening questionnaire. The researcher did not find strong correlations among the AIM scales with habitual responsiveness nor did the AIM scales reliably predict situational responsiveness. However, Lehmann did notice, “Music majors reported significantly higher responsiveness than non-musicians” (p. 84). The researcher argued, “that reported affective response in music is partially mediated by expertise-related mechanisms” (p. 84). This study supports the idea that music training influences individuals’ emotional response to music. The present researcher is curious to know if the amount or type of music training would influence adjudicators’ ratings of musical performance.

Tsay (2013) determined there were no significant differences between non-musicians and expert musicians when determining winners of a music competition. Tsay stated, “. . . neither musical novices nor professional musicians were able to identify the winners based on sound recordings or recordings with both video and sound” (p. 14580). Tsay concluded that music training is unnecessary for music competition.

Researchers sought to reveal characteristics related to musicians and non-musicians’ differing perceptions of a musical performance and discussed the benefits of training in the forms of music experience, instruction, teaching level, and adjudicator training. Trained and untrained musicians, as well as non-musicians, seemed to hear and

assess music performance differently. Although there was no difference among musicians and non-musicians in their judgments about music competition winners (Tsay, 2013), their opinions differed among specific details about performance. People's expectations about a performance (i.e., as affected by prior experience) influenced their evaluation of the performance (Brown and Novak, 2007; Kinney, 2009; Mills, 1991). It seems that in choral music performance settings, all people who are engaged in the activity perceive musical performance from multiple perspectives. The present researcher would like to contribute to this area of knowledge by investigating adjudicators' academic training. Are there differences among adjudicators regarding influences on their rating decision of choral music performance by the types of academic degrees they hold?

### **Selection of Evaluative Criteria and Evaluative Instruments**

#### **Evaluation Criteria**

Adjudication ratings forms are designed to provide two types of information: (a) numerical ratings representing overall evaluations of the musical performance and (b) comments to communicate the rationale for the rating. Most researchers who explored music performance evaluation investigated the measure, or tool, used to rate a musical performance. Prior researchers in the area of music performance adjudication provided evidence that music performance assessment rubrics did not provide strong inter-rater reliability (Bergee, 2003; Cooksey, 1977; Hash, 2012; Latimer, 2007; Latimer et al., 2010; Norris & Borst, 2007). In these studies, researchers used a wide range of participants to test hypotheses regarding the inter-rater reliability of each measure.

Researchers sought to know if adjudicators agreed with each other on the performance rating measures. The range of participants included highly respected and experienced musicians', less-experienced musicians', and non-musicians' evaluations of music performances. In most cases, the inter-rater reliability, or the agreement among adjudicators on music performance characteristics, was not acceptable.

Music performance ratings forms vary from open-ended and seemingly blank sheets of paper to highly structured rubrics with lengthy descriptions of every possible aspect of a musical performance. A general understanding among adjudicators is that open-ended tools are considered 'global' or 'holistic' forms utilizing a holistic approach while standardized forms with categories and definitions are 'specifics' forms (Cooksey, 1977; Mills, 1991; Radocy, 1989; Stanley, Brooker, & Gilbert, 2002). Global forms allowed adjudicators to evaluate a performance holistically and assign a final rating that they felt best represented the performance as a whole. Specifics (or segmented) forms required adjudicators to measure observable musical aspects of a performance, with each aspect receiving a weighted score and summed with the other aspect scores to produce an overall score.

Cooksey's (1977) study is considered to be seminal work in the field regarding choral ensemble adjudication. The purpose of this study was to construct and test a rating scale for the evaluation of high school choral music performance, using a facet-factorial approach. This research study sought to develop a new means to evaluate performance achievement. The researcher used many methods to gather information to create the new measurement tool. After the researcher created the new measurement tool, it was field

tested in a choral adjudication setting, and then the data were analyzed using statistical analysis. The selection and identification of factors that formed the basis structure of the choral performance rating scale resulted from both statistical and musical considerations.

Cooksey's (1977) study comprised many parts and was highly complex in its design. First, the researcher sought to know the current relevant criterion for high school choral performance achievement. Second, he sought to create a new ratings scale rubric that would result in higher inter-judge reliability and criterion-related validity than the nationally accepted choral performance rubric. Third, after he created a new rating scale rubric, he tested it and modified it. Lastly, he compared the new rating scale rubric to the accepted national rating scale rubric. Seven factors of choral performance were produced by the analysis: diction, precision, dynamics, tone control, tempo, balance and blend, and interpretation and musical effect. Thirty-six items-dimensions were selected to form the subscales to measure the seven factors.

Jones (1986) developed a rating scale for high school vocal solo performance evaluation. The researcher argued that music performance evaluation was subjective in nature, "The assessment of achievement of complex behaviors such as music performance is generally based on subjective observations that are uncontrolled and unstructured" (p. ix). He collected information descriptive of high school vocal solo performance from available vocal pedagogy literature and statements from members of National Association of Teachers of Singing (NATS). The review of literature and NATS members' statements produced a pool of 168-items. These items were paired with a five-point Likert-type scale and then used by voice teachers to rate thirty high school



singers. The researcher used factor analysis to determine the foundation and structure of vocal performance and assessment. The study produced a 32-item scale based on a five-factor structure of vocal performance called the Vocal Performance Rating Scale (VPRS). Five factors of vocal performance were: Interpretation/musical effect, tone/musicianship, technique, suitability/ensemble, and diction.

Researchers in other music performance domains generated similar music performance ratings measures (Mills, 1991; Smith & Barnes, 2007; Zdzinki & Barnes, 2002). Zdzinki and Barnes's (2002) study resulted in a measure for string instrument performance. The measure consisted of five factors (Interpretation/musical effect, articulation/tone, intonation, rhythm/tempo, and vibrato) with 23 subscales producing reliability ranging from .873 to .936. Smith and Barnes (2007) sought to develop and validate a factor-derived measure for orchestra ensemble performance achievement. They assembled a pool of statements used in evaluating orchestra performance at the middle school and high school level, then paired the statements with a nine-point Likert-type scale, and asked experts to evaluate orchestras using the measure. "Two rounds of validation showed high correlations with MENC's [Music Educators National Conference] adjudication form and a ranking task; the initial factor structure was not duplicated exactly, indicating directions for future research" (p. 268).

Mills (1991) explored holistic (global) and segmented assessment in assessing music performance, as they were used in summative assessments of music programs at the end of the school year. Mills sought to compare constructs from holistic assessment with the approved standardized segmented assessment method. First, the researcher

determined the constructs used in holistic music assessment. Then she used those constructs to generate a segmented assessment. Finally, she compared the segmented assessment to the approved Associated Board of the Royal Schools of Music (ABRSM) measure.

Mills' (1991) study was divided into two stages. In stage one, Mills asked music teachers to select their top soloist performers to provide the stimulus for the experiment. Eight students were video-recorded performing a solo in their primary instrument. Mills then asked 11 student teachers to view these videos and rank them according to overall performance quality, and then each student teacher completed an interview with the researcher. Next, Mills selected three videos from the ranked performances. Mills had the student teachers watch the top three selected videos and asked each of them a series of questions. Student teachers' responses used to generate a list of constructs. In stage two, Mills had 29 participants evaluate 10 video-recorded performances using both the approved ABRSM measure and the researcher-constructed measure. After the evaluation, participants' scores were converted and analyzed statically. Mills stated:

Unremarkably, a correlation analysis shows a positive correlation between each of the constructs and the overall rank ranging from  $r = 0.4$  (C6) to  $r = 0.7$  (C10 and C11) ( $n = 290$ ). Multiple regression analysis with the rank as the dependent variable and the scores on the constructs as the independent variables resulted in the production of a regression equation which accounts for 71% ( $n = 290$ ) of the variance in the ranks. (p. 179)

Mills argued that holistic assessment has advantages over segmented assessment. The researcher sought to provide evidence that one method was more useful than the other. In her discussion, Mills stated that holistic assessment is "more musically credible, in the

sense that it is more like assessment made of musical performance in the real world” (p. 179.) Mills’ research seemed to not support the concept of using criterion-based multi-dimensional rubrics. Mills seemed to support the traditional holistic-global approach to assessing music.

Wrigley and Emmerson (2011) investigated ways to improve the quality of music performance assessment at the tertiary level. The study was a way for researchers to respond to the accountability imperative in higher education music programs. “An enhanced scientific methodology was employed incorporating ecological validity and using recognized qualitative methods involving grounded theory and quantitative methods involving confirmatory factor analyses” (p. 97). Phase one of the qualitative analysis revealed many emergent themes. Further analysis of these themes produced lists (for each instrument family) of 40-50 terms reflective of musical performance constructs. In phase two, researchers refined item lists to construct ratings scales for each instrument family. In phase three, a Performance Evaluation Report Scale (PERS) was constructed and implemented. Validity and internal reliability of the PERS were calculated and determined in phase four, which resulted in the refinement of the measure. Results from the study indicated that examiners in each instrument domain of classical music performance used between 15 and 17 constructs with associated descriptors. In addition, researchers found “at least two dimensions referring to technical proficient and another relating to musicality and interpretation” (p. 97).

## Reliability

Bergee (2003) examined the inter-judge reliability of university music faculty members' evaluations of music students' end-of-semester jury performances. Twenty-five applied music performance faculty evaluated end-of-semester juries of applied music performance students at a large university. Participants (faculty instructors) represented various music performance domains: brass ( $n = 4$ ), percussion ( $n = 2$ ), woodwind ( $n = 5$ ), voice ( $n = 5$ ), piano ( $n = 3$ ), and strings ( $n = 5$ ). Participants completed a criterion-specific ratings form for each jury performance within their respective music performance domain. Following the completion of the ratings form, participants assigned each performer a global letter grade—a letter grade that represented the performers entire musical performance. Evaluators' global letter grade scores, total ratings scale scores, and sub-scale scores were examined using quantitative statistics. Inter-judge reliability within each domain was calculated. "All possible permutations of two, three, and four [evaluators] were examined for inter-judge reliability, and averaged correlations, standard deviations, and ranges were determined" (p. 137). Averaged inter-judge reliability results were consistently good regardless of adjudicator panel size (i.e., brass = .82, percussion = .93, woodwind = .83, voice = .81, piano = .82). Panel sizes ranged from two (percussion,  $n = 2$ ) to five (voice, woodwind, strings,  $n = 5$  each). In larger adjudicator panels ( $n = 4$  and 5), "rating scale total score reliability was consistently but not greatly higher than reliability for the letter-grade assessment" (p. 137). Permutations of the smaller number adjudicator panels ( $n = 2$  and 3) exhibited greater variability, range, and less uniformity than larger adjudicator panels. Bergee suggested that adjudicator panels at music

performance assessments consist of a minimum of five adjudicators in the context of end-of-semester jury performances.

Latimer (2007) investigated adjudicator reliability by comparing utilization of an authentic State Festival Choir audition form to utilization of a global score form. The researcher selected thirty participants (i.e., ten experienced adjudicators, ten music educators without adjudication experience, and ten non-music educators without adjudication experience), and used both the Kansas State Music Educators Association audition form and the American Choral Directors Association audition form to evaluate fifteen female high school singers. Results from the study confirmed prior research in the domain of vocal music evaluation. Adjudicators' scores were not reliable in both inter-adjudicator and intra-adjudicator reliability. Latimer (2007) concluded that the Kansas State Music Educators Association adjudication form needed to be re-designed.

Hash (2012) examined procedures for analyzing ratings of large ensemble high school band music performance evaluations and provided data with which to compare results from similar events. The researcher pursued three research questions: (a) what was the distribution of ratings among bands, (b) what was the reliability of individual judging panels, and (c) did average final ratings differ among judges, contest sites, years, or classifications? The researcher used several statistical tests (i.e., Spearman correlation, Cronbach's alpha, Fischer's  $z$  transformation, Friedman and Kruskal-Wallis ANOVAs, and Mann-Whitney  $U$  tests). Most high school bands received Superior or Excellent ratings. Individual judges' scores for the concert portion reached a moderately high level of agreement ( $r_s = .75$ ), but did not achieve the minimum benchmark of .80. "Results of

Friedman ANOVAs revealed significant differences among individual judges' ratings within eight of the eighteen adjudicator panels indicating that some evaluators graded at a higher degree of severity than others did" (p. 90).

Norris and Borst (2007) compared the reliability of a common school choral festival adjudication form with that of a second form that was a more descriptive extension of the first. Norris and Borst sought to learn if there was a difference between two music performance rubrics: (a) the accepted standard form already in use and (b) a new and more detailed form created by the researchers. The researchers gathered four highly qualified adjudicators and instructed them to adjudicate two recorded choral performances using the two different performance rubrics. The researchers discovered no significant differences between forms for specified categories of tone, diction, blend, intonation, rhythm, balance, total score, and overall rating. The researchers used paired-sample *t*-tests for their analysis of adjudicator scores. They completed many *t*-tests, increasing the probability of a Type I error, possibly rejecting the null hypothesis when it might be true. An ANOVA might have been a more efficient test rather than computing so many *t*-tests. When performing an ANOVA a researcher can test for other possible assumptions related to variability between groups of scores, providing more depth and understanding.

Latimer et al. (2010) investigated the reliability and perceived pedagogical utility of a multidimensional weighted performance assessment rubric for high school large ensemble festivals. The researchers gathered 2,016 adjudicator rubrics and completed questionnaires from adjudicators and ensemble directors ( $n = 515$ ). "Dimension

reliability ranged from moderately low ( $W = .47$ ) to moderate ( $W = .77$ ). Total score reliability was moderately high ( $W = .80$ ) and rating reliability was moderate ( $W = .72$ )” (p. 168). One of the primary goals of this study was to test the reliability of the rubric. Investigators determined that the rubrics used for high school large ensemble performance were moderately good and consistent with prior researchers’ determinations (Ciorba & Smith, 2009; Saunders & Holahan, 1997).

Prior research in the domains of evaluative criteria and instruments provides rich information regarding the constructs of music performance, ratings form and rubric construction, and reliability of the instruments as well as adjudicators. Several of these studies obtained their data from literature, recorded performances, adjudicators’ completed ratings forms, and questionnaires. This domain of research has been investigated heavily, however; there is limited study regarding discrete aspects of influences on adjudicators’ rating decisions of music performance (McPherson & Thompson, 1998).

### **Modified Delphi Method**

The first use of the Delphi Method was Dalkey and Helmer (1963). They sought to obtain the most reliable consensus from a group of experts. The researchers created the method during the cold war to determine the number of A-bombs needed for strategic military use. Dalkey and Helmer (1963) stated that the experiment applied “expert opinion to the selection, from the viewpoint of a Soviet strategic planner, of an optimal U. S. industrial target system and to the estimation of the number of A-bombs required to reduce the munitions output by a prescribed amount” (p. 458).

The original Delphi method involved the use of repeated questioning of seven experts. “Four of these [participants] were economists, one was a physical-vulnerability specialist, one a systems analyst, and one an electronics engineer” (Dalkey & Helmer, 1963, p. 460). There was a total of five questionnaires sent to participants; one questionnaire a week for five weeks. Interviews were held after the first and third questionnaires. Participants’ answers were in the form of numerical data. These data were summed and averaged for each participant. The final average was used as a consensus of estimates (Dalkey & Helmer, 1963, p. 465). The method was used to predict future incidences (Geisler, 2014).

Since Dalkey and Helmer (1963), other researchers used a modified form of the Delphi method in their investigations (Haire, 2015; Geisler, 2014; Gous-Kemp, 2014). Gous-Kemp (2014) described the process of a modified Delphi method when seeking a consensus among a group of experts:

Experts were selected from different appropriate fields such as music education, inclusive education, creativity and materials development. An open-ended questionnaire was developed and e-mailed to experts, together with the first draft of the program. Responses were analyzed, categorized and incorporated into the program. Only one round was necessary. (p. 235)

Haire (2015) utilized a Delphi-like methodology to investigate reasons why male chorus students signed up for choir at their school. The study was Delphi-like in that he used one open-ended question to gather information from participants. Participants were instructed to be as detailed as possible when responding to the open-ended question (Haire, 2015). Geisler (2014) “utilized a modified Delphi method to investigate the



opinions and perspectives of experts in the field of music education in the State of Ohio” (p. 2). In his study, Geisler suggested that the modified Delphi technique was effective because of its “flexibility of administration, the ability to provide accurate expert perspectives on current issues that directly affect the field of music education, and, the anonymity of participants” (p. 26).

### **Summary of Related Literature**

The present research study utilized McPherson and Thompson’s (1998) PMAMP as a fundamental structure for inquiry into the assessment of musical performance. Music performance assessment is pervasive in research, school music programs, and in competitions and festivals. Music performance assessment exists within many arenas: from instrumental music to vocal music, solo-based performance to ensemble-based performance, primary/secondary school music to collegiate music programs, auditions for prestigious musical organizations and ensemble membership, and so on. Music performance assessment consists of many complex interacting factors. McPherson and Thomson’s PMAMP aided the present researcher in conducting a thorough review of available research. The assessment of music performance has many implications and deserves continued research.

When investigating music performance adjudication, one must consider the purposes and context of the adjudication (McPherson & Thompson, 1998). Stakeholders (including researchers, school administrators, adjudicators, event organizers, etc.) of MPA should understand the motivations of choral directors and students in school choral programs to participate in MPA (Millard, 2014). Adjudicators should understand

philosophical underpinnings and reasons for the adjudication event. Otherwise, without this knowledge, adjudicators might rate the quality of performance based solely on personal preference and experience (Ekholm, 2000; Ford, 1999), and will be affected by bias conditions (Cassidy & Sims, 1991; Radocy, 1976), regardless of ratings rubrics or standardized assessment forms (Cassidy & Sims, 1991; Ekholm, 2000; Ford, 1999; Radocy, 1976).

Musical and nonmusical factors influenced both the performance and adjudicators' ratings of musical performance (McPherson & Thompson, 1998). As described by Landy and Farr (1980), music performance assessment is a constellation of physical and mental operations in isolation. The adjudicator, performer, ratings instrument, evaluation process, and information processing coexist and interact to produce a final result of a performance. Performance proportions (Bergee & Platt, 2003; Bergee & Westfall, 2005), contextual limitations (Bergee & Platt; 2003, Welch, 1994), visual aspects of the performers (Mitchell & MacDonald, 2016; Tsay, 2013), conductor's appearance (Madsen, 2009; Morrison et al., 2009; Napoles, 2013), ensemble movement (Junchniewicz, 2008, Vasconcellos, 2002), ethnicity (Brown & Novak, 2007; Cheek, 2007; Elliot, 1995/1996; Vanweelden & McGee, 2007), performers' first impression (Platz & Kopiez, 2013), performance attire and stage deportment (Howard, 2012), and other non-musical factors have been found to have significant effects on adjudicators' perceptions and consequently their ratings of performance quality. Furthermore, evaluators' personality influenced how they perceived music as well as how they reacted emotionally and physically to it (Colver & El-Alayi, 2016; Thomas et al., 2015).

In general, research supports the notion that music experience influences adjudicators' ratings of music performance (Kinney, 2009; Mills, 1991). This research focused on music experience as: (a) instruction in music classes (Hair, 1981), (b) teacher-career level (Hewitt & Smith, 2004), and (c) academic-teaching level (Howard, 2012). The present researcher recognizes an area of possible importance that is not represented in available research, i.e., adjudicators' adjudication experience, specifically. The present research study investigated adjudicators' years of adjudication experience as a factor of influence on ratings of choral ensemble performance quality.

The body of research in music performance assessment presented a number of studies related to instrumental performance (Bergee, 2003; Bergee, 2007; Bergee & Platt, 2003; Bergee & Westfall, 2005; Burns, Vernon, & Hinkle, 1985; Ciorba & Smith, 2009; Geringer & Madsen, 1998; Hash, 2012; Hewitt & Smith, 2004; Juchniewicz, 2008; Krueger, 1966; Landy & Farr, 1980; Madsen et al., 2007; McPherson, 1995; McPherson, 1993; McPherson & Thompson, 1998; Mitchell & MacDonald, 2016; Platz & Kopiez, 2013; Saunders & Holahan, 1997; Silveira, 2014; Silvey & Risher, 2015; Smith & Barnes, 2007; Springer & Schlegel, 2016; Stanley et al., 2002; Wrigley & Emmerson, 2011; Zdzinski & Barnes, 2002) and solo-vocal performance (Bergee, 2003; Bergee & Platt, 2003; Bergee & McWhirter, 2005; Ciorba & Smith, 2009; Geringer & Madsen, 1998; Howard, 2012; Jones, 1986; Stanley et al., 2002). Many researchers focused on the selection of evaluative criteria (Cooksey, 1977; Jones, 1986; McPherson, 1995; Smith & Barnes, 2007; Stanley et al., 2002; Wrigley & Emmerson, 2011) and the evaluation of measurement instruments (Bergee, 2003; Bergee & Platt, 2003; Ciorba & Smith, 2009;

Cooksey, 1977; Kinney, 2009; Latimer et al., 2010; McPherson, 1995; Mills, 1991; Norris & Borst, 2007; Saunders & Holahan, 1997). Another area of focus in research was training (Hewitt & Smith, 2004; Ward, 2013); however, the researchers reviewed training only in the forms of experience (Brown and Novak, 2007; Kinney, 2009; Mills, 1991), adjudication training (Ward, 2013), and teaching level (Hewitt & Smith, 2004). The present researcher was unable to locate research that investigated adjudicators' academic level (i.e., highest degree earned and type of degree) as a factor of influence on ratings of choral ensemble performance quality.

There appeared to be limited research in the assessment of choral ensemble performance (Cassidy & Sims, 1991; Cooksey, 1977; Daugherty 1999; Ekholm, 2000; Grant & Norris, 1998; Latimer, 2007; Latimer et al., 2010; Madsen, 2009; Millard, 2014; Norris & Borst, 2007; Robinson, 1988; Selvey, 2014; Stutheit, 1994; Vasconcellos, 2002; Ward, 2013, Welch, 1994) compared to the instrumental field. More research in the domain of choral ensemble assessment is needed in order to further understand the assessment of choral music performance.

### **Research Questions**

1. What factors influence adjudicators' decisions when adjudicating choirs?
2. Do differences exist among adjudicators' influences on rating decisions by years of adjudication experience?
3. Do differences exist among adjudicators' influences on rating decisions by academic training?

## **CHAPTER III**

### **METHOD**

#### **Restatement of the Purpose**

The purpose of the present study was to determine what influenced choral music adjudicators' rating decisions about the quality of a choral ensemble's performance. The present research study investigated what influenced choral performance adjudicators' ratings and explored possible differences among adjudicators regarding those influences. Music performance adjudication is a common experience for choral teachers and students, however music educators have expressed concerns regarding adjudicator bias (Corbin, 1995; Fox, 1990; McPherson & Thompson, 1998; Robinson, 1990; Ward, 2013). By examining aspects that influence choral music performance adjudicators' rating decisions of a choral performance, the researcher hopes to learn more about what influences adjudicators' judgments.

#### **Research Questions**

1. What factors influence adjudicators' decisions when adjudicating choirs?
2. Do differences exist among adjudicators' influences on rating decisions by years of adjudication experience?
3. Do differences exist among adjudicators' influences on rating decisions by academic training?

## Participants

Participants for the present study ( $N = 71$ ) included individuals who serve as choral performance adjudicators within the Southern Division of the National Association for Music Education (NAfME) and the American Choral Directors Association (ACDA). The Southern Division of NAfME and ACDA include the states of Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and West Virginia, and the Commonwealths of Kentucky and Virginia (National Association for Music Education, December 2015). Demographic information was gathered for each participant, including (a) age, (b) gender, (c) years of teaching experience (d) years of adjudication experience, (e) current school location (i.e., urban, suburban, rural), (f) primary teaching level (i.e., primary, secondary, collegiate, retired), and (g) education background (i.e., degree levels and focus). Participants were recruited with the assistance of the choral activities chairpersons from each of the eleven music educator associations within the Southern Division of the NAfME and ACDA.

Forty-four percent of the participants were male. Fifty-six percent of the participants were female (see Table 1). Among the participants, 1.4% was between the ages of 26 and 30, 15.5% were between the ages of 31 and 35, 7% were between the ages of 36 and 40, 11.3% were between the ages of 41 and 45, 9.9% were between the ages of 46 and 50, 8.5% were between the ages of 51 and 55, 16.9% were between the ages of 56 and 60, 19.7% were between the ages of 61 and 65, 5.6% were between the ages of 66 and 70, and 4.2% were between the ages of 76 and 80 (see Table 2).

Table 1

## Participants' Gender

Gender	<i>n</i>	%	Cumulative Percent
Male	31	44%	44%
Female	40	56%	100%
Total	71	100%	

Table 2

## Participants' Age

Age	<i>n</i>	%	Cumulative Percent
26-30	1	1.4%	1.4%
31-35	11	15.5%	16.9%
36-40	5	7.0%	23.9%
41-45	8	11.3%	35.2%
46-50	7	9.9%	45.1%
51-55	6	8.5%	53.5%
56-60	12	16.9%	70.4%
60-65	14	19.7%	90.1%
66-70	4	5.6%	95.8%
76-80	3	4.2%	100.0%
Total	71	100.0%	

Among the participants, 1.4% had zero to five years of choral teaching experience, 8.5% had six to ten years of choral teaching experience, 14.1% had 11 to 15 years of choral teaching experience, 12.7% had 16 to 20 years of choral teaching experience, 9.9% had 21 to 25 years of choral teaching experience, 11.3% had 26 to 30

years of choral teaching experience, and 42.3% had 30 or more years of choral teaching experience (see Table 3).

Table 3

Participants' Years of Choral Teaching Experience

Years of Choral Teaching Experience	<i>n</i>	%	Cumulative Percent
0-5	1	1.4%	1.4%
6-10	6	8.5%	9.9%
11-15	10	14.1%	23.9%
16-20	9	12.7%	36.6%
21-25	7	9.9%	46.5%
26-30	8	11.3%	57.7%
30 or more	30	42.3%	100.0%
Total	71	100.0%	

Among the participants 26.8% worked in urban schools, 45.1% worked in suburban schools, and 26.8% worked in rural schools (see Table 4). The present researcher did not define school locations in the data collection instrument. Participants self-selected their school's location. Regarding participants' priority school-level assignment, 1.4% were at the primary-school level (kindergarten through grade 5), 2.8% were at the middle school level, 59.2% were at the high school level, and 35.2% were at the college and university level (see Table 5).



Table 4

## Participants' School Location

School Location	<i>n</i>	%	Cumulative Percent
Urban	19	26.8%	27.1%
Suburban	32	45.1%	72.9%
Rural	19	26.8%	98.6%
Missing	1	1.4%	100.0%
Total	71	100.0%	

Table 5

## Participants' School-Level Assignment

School-Level Assignment	<i>n</i>	%	Cumulative Percent
Elementary School	1	1.4%	1.4%
Middle School	2	2.8%	4.3%
High School	42	59.2%	64.3%
College and University	25	35.2%	98.6%
Missing	1	1.4%	100.0%
Total	71	100.0%	

### Data Collection Instrument

The data collection instrument was designed using a three-step process. In Step 1, the researcher created an open-ended questionnaire (Appendix A) that included one item: "What aspects do you consider when you adjudicate a choir performance at your state's music performance adjudication event? Please include as many aspects as you can think

of in your response” and administered that questionnaire to a randomly selected group of eight individuals drawn from the population of choral adjudicators in the state of North Carolina. Choral adjudicators’ responses were reviewed and a list of discrete performance aspects was generated from those responses (Appendix B). In Step 2, a list of performance aspects was generated from aspects found in the related literature (Appendix C.1) and merged with the combination of adjudicators’ responses (Appendix C.2) (Haire, 2015). A discrete list of performance aspects was generated by combining similar items, revealed via adjudicators’ responses and the items described in the related literature (Appendix C.3). In Step 3, the compiled list of items was used to generate a pilot survey (Appendix D), which was sent to a panel of experts (i.e., choral activities chairpersons of each of the music educators associations within the Southern Divisions of NAfME and ACDA) for review. The experts reviewed the pilot survey and made suggestions to add, remove, or edit items. Information generated from the experts’ feedback was used to construct a closed-ended survey that was used to collect data for the present study.

The researcher developed the *Choral Adjudicator Preference Scale (CAPS)* to measure the importance of each aspect vetted via the open-ended questionnaire, the related literature, and the choral adjudicator experts (Appendix E). For each aspect, prospective participants were asked to indicate, using a five-point Likert-type scale, how strong of an influence that item had on their rating decision when evaluating choral performances (i.e., 5 = Extremely Strong, 4 = Very Strong, 3 = Strong, 2 = Somewhat Strong, 1 = Not Strong). The instrument included a demographic section that requested

information including the participant's (a) age, (b) gender, (c) years of adjudication experience, (d) education background (i.e., degree levels and focus), (e) current school location (i.e., urban, suburban, rural), and (f) primary teaching level (i.e., primary, secondary, collegiate, retired) (Wagoner, 2011).

Reliability of the CAPS was calculated using a Cronbach's Alpha statistical procedure to measure internal consistency. Validity of that CAPS was ensured through the three-step instrument-construction process.

### **Data Collection Procedures**

Approval for the present study was granted by the Institutional Review Board (IRB) at the University of North Carolina at Greensboro (UNCG) (Appendix F). The NCMEA Choral Activities Chair was contacted to secure a list of current choral music performance adjudicators. A number was assigned to each adjudicator and a random number generator was used to select eight individuals; ensuring appropriate variability existed across all levels of each demographic variable. Each potential participant was sent a recruitment letter that requested they participate in the study (Appendix G). To those who agreed to participate, the researcher sent the open-ended questionnaire for completion. In the circumstance that adjudicators chose not to participate, those names were removed from the list of potential participants. Remaining adjudicators on the list were randomized again, and potential participants were sent a recruitment letter. A total of 29 adjudicators were asked to participate in the first part of the study. A total of eight participants agreed to participate and received the questionnaire for the initial part of the study. The open-ended questionnaire was sent to participants on February 11, 2016, and

requested that completed open-ended questionnaires be returned on or before April 24, 2016. Once all open-ended questionnaires were received, a list of discrete items and a frequency count of each aspect described was generated from their statements.

After the list of discrete items from adjudicators' statements was generated, those items were combined with discrete items found in the related literature. The researcher generated a pilot survey from those discrete items. On September 7, 2016, experts in the field of choral music adjudication (i.e., Choral Activities Coordinators within the NAFME Southern Division) were asked to review the pilot survey and make suggestions for improvement—respond with suggestions to add, remove, or edit items.

From experts' feedback, a close-ended data-gathering instrument (CAPS) was developed for the data gathering procedure. After all items on the CAPS were finalized, a link for the online survey was sent to all potential participants (i.e., choral performance adjudicators within the Southern Division of NAFME and ACDA).

Qualtrics, an online-based survey software program, was programmed to contact participants using e-mail addresses provided by the Southern Division choral activities chairs and published online adjudicator lists. A recruitment letter (Appendix H) was sent to each choral adjudicator. If a choral adjudicator decided to participate he or she was directed immediately, via emailed invitation with a link, to the research study's consent form (Appendix I) and then to the survey. The survey was made available the day the recruitment letter was sent. Qualtrics was programmed to send periodic reminders to potential participants to complete the CAPS. Data were collected from participants beginning on October 1, 2016 and ending November 1, 2016.

### **Data Analysis Procedures**

Once data were collected using the CAPS, all data were analyzed using descriptive statistics, Cronbach's Alpha, factor analysis, and analysis of variance (ANOVA). A Cronbach's Alpha statistical procedure was calculated to determine reliability of the close-ended data-gathering instrument. IBM's SPSS Statistics version 24 was used to calculate Cronbach's Alpha. Cronbach's alpha ( $\alpha$ ) was .934, which indicated a high level of internal consistency. Validity for the CAPS was established through three sources, (a) an open-ended questionnaire sent to the eight choral adjudicators, (b) a thorough review of the related literature and (c) verification by the choral activities chairpersons of the eleven music educators associations of the NAFME and ACDA Southern Division.

For research question one (i.e., "What factors influence adjudicators' decisions when adjudicating choirs?"), a factor analysis procedure was calculated to determine whether aspects choral adjudicators identified as influencing their decisions could be reduced to a smaller number of discrete factors. The goal of this analysis was to determine the most salient aspects of influences on choral adjudicators' rating decisions of choral performance.

For research question two (i.e., "Do differences exist among adjudicators' influences on rating decisions by years of adjudication experience?") a two-way repeated measures ANOVA was used to analyze whether differences exist across the discrete factors found in question one by participants' years of adjudication experience (0-5, 6-10, 11-15, 16-20, 21-25, 26-30, 31+).

For research question three (i.e., “Do differences exist among adjudicators’ influences on rating decisions by academic training?”) a two-way repeated measures ANOVA was used to analyze whether differences exist across the discrete factors found in question one by participants’ academic training as described in the demographic section of the CAPS survey (degree level and focus).

## CHAPTER IV

### RESULTS

#### Review of Research Questions

Three research questions were formulated for the present study:

1. What factors influence adjudicators' decisions when adjudicating choirs?
2. Do differences exist in adjudicators' influences on rating decisions by years of adjudication experience?
3. Do differences exist in adjudicators' influences on rating decisions by academic training?

The sequence of research questions, one through three, structures the rest of Chapter IV.

#### **Research Question 1: Factors Influencing Choral Adjudicators' Rating Decisions**

Data gathered from the *Choral Adjudicator Preference Scale (CAPS)* were used to answer research question one: "What factors influence adjudicators' decisions when adjudicating choirs?" IBM's SPSS software was employed to calculate descriptive statistics for each item including rank order by mean score, the mean, standard deviation, and range. Microsoft Excel 2011 was used to organize data calculated by SPSS.

#### ***Choral Adjudicator Preference Scale Descriptive Statistics***

Item responses on the *Choral Adjudicator Preference Scale* were structured in the form of a Likert-type scale that asked participants to describe how strong of an influence

each item had on their rating decisions when evaluating a choral music performance (i.e., 5 = Extremely Strong, 4 = Very Strong, 3 = Strong, 2 = Somewhat Strong, 1 = Not Strong). Using these numerical values, item rank order by mean score, mean, standard deviation, and range were calculated using IBS's SPSS software.

The highest item response mean of the *Choral Adjudicator Preference Scale* was 4.83 for the item "Accuracy of pitches." This high mean response revealed that a choral ensemble's performance of accurate pitches from a musical selection was a very strong to extremely strong influence on many choral adjudicators' ratings of choral performance. The lowest item mean response was 1.25 for the item "Prior knowledge regarding the performers." This low mean revealed that, for many choral adjudicators, this item was not an influence on their choral performance rating decision.

Results from analyzing the mean score, standard deviation, and range of all item responses from the *Choral Adjudicator Preference Scale* are included in Table 6. These results were used to determine the importance of influence of each item on choral adjudicators' ratings of choral performance as well as how participants were clustered around the means.



Table 6

*Choral Adjudicator Preference Scale: Item Descriptions and Descriptive Statistics, Sorted by Item Mean*

Item Ranked by Mean	Item Description	Item Mean	Item Standard Deviation	Item Range	Minimum	Maximum
1	Accuracy of pitches	4.83	.41	2	3	5
2	Tone quality	4.75	.44	1	4	5
3	Accuracy of rhythm	4.72	.51	2	3	5
4	Melodic intonation	4.69	.55	2	3	5
5	Appropriateness of vowels	4.65	.61	2	3	5
6	Harmonic intonation	4.64	.54	2	3	5
7	Diction	4.55	.60	2	3	5
8	Musical expression	4.48	.65	2	3	5
9	Sense of ensemble	4.41	.71	2	3	5
10	Vocal technique	4.38	.70	3	2	5
11	Blend	4.33	.72	3	2	5
12	Phrasing	4.30	.64	2	3	5
13	Breath support	4.27	.68	2	3	5
14	Musicianship of the ensemble	4.23	.78	3	2	5
15	Balance	4.21	.74	3	2	5
16	Precision	4.18	.68	3	2	5
17	Appropriateness of consonants	4.16	.81	3	2	5
18	Tone control	4.14	.75	3	2	5
19	Dynamics	4.01	.71	2	3	5
20	Appropriateness of music for the ensemble	4.00	.88	3	2	5
21	Adherence to musical instructions notated on musical score	3.94	.92	3	2	5
22	Ensemble's commitment to a musical performance	3.85	1.10	4	1	5
23	Interpretation	3.82	.82	3	2	5
24	Adherence to standard performance practice of the musical selection	3.76	.84	3	2	5
25	Ensemble's responsiveness to the conductor	3.68	.92	4	1	5
26	Musical style	3.66	.93	4	1	5
27	Appropriateness of tempo	3.61	.90	4	1	5

Table 6

Cont.

Item Ranked by Mean	Item Description	Item Mean	Item Standard Deviation	Item Range	Minimum	Maximum
28	The degree to which the accompanist and the ensemble work together to enhance the musical performance	3.46	1.07	4	1	5
29	Adherence to rules of conduct and behavior for the event	3.42	1.33	4	1	5
30	Body alignment	3.27	.88	4	1	5
31	Appropriateness of vibrato	3.20	1.02	4	1	5
32	Memorization	3.11	1.32	4	1	5
33	Difficulty level of music	2.82	1.19	4	1	5
34	Rubato	2.75	1.01	4	1	5
35	Expressiveness of the conductor	2.72	.99	4	1	5
36	Appearance	2.69	.96	4	1	5
37	How the ensemble walks onto the stage	2.63	1.17	4	1	5
38	Facial expression of the Ensemble	2.62	.88	4	1	5
39	Method used to evaluate music performance (Multidimensional rubric, holistic, etc.)	2.59	1.19	4	1	5
40	How the ensemble walks off the stage	2.39	1.04	4	1	5
41	Ensemble movement	2.37	.96	3	1	4
42	First impression	2.28	1.00	4	1	5
43	Age level	2.25	1.38	4	1	5
44	Conducting gesture	2.24	.93	3	1	4
45	Context of the performance event (purpose and location)	2.21	1.21	4	1	5
46	Standing formation of the group	2.13	1.01	3	1	4
47	Knowledge regarding the amount of time the choral ensemble rehearsed prior to the performance	1.97	1.06	4	1	5
48	Performance space	1.63	.91	3	1	4
49	Ensemble size	1.42	.73	3	1	4
50	Time of day	1.34	.65	3	1	4

Table 6

Cont.

Item Ranked by Mean	Item Description	Item Mean	Item Standard Deviation	Item Range	Minimum	Maximum
51	School size	1.34	.70	3	1	4
52	Audience response	1.31	.71	3	1	4
53	Prior knowledge regarding the performers	1.25	.55	2	1	3

### ***Choral Adjudicator Preference Scale Principal Component Analysis (PCA)***

To further answer Research Question 1, “What factors influence adjudicators’ decisions when adjudicating choirs?” a principal component analysis (PCA) was calculated to determine whether groups of items that choral adjudicators identified as influencing their decisions would coalesce under a smaller number of discrete factors (Haire, 2015; Teachout, 2004, 2008). Principal component analysis is a form of factor analysis that focuses on variable or item reduction. The goal of the present analysis was to uncover the most salient aspects of choral adjudicator participants’ self-reported influences on their rating decisions of choral performance quality. *Laerd Statistics* (2015a) was employed as a guide for using IBM’s SPSS software to run a PCA and determine (a) the number of discrete factors, (b) use and type of rotation technique, (c) total variance explained, and (d) interpretation of the results.

To determine if a PCA would be an appropriate method to reduce the items on the *Choral Adjudicator Preference Scale (CAPS)* to a more discrete number of factors, and reveal possible underlying latent patterns, the data were tested to determine if they passed

the assumptions for linearity and sampling adequacy. Because a few choral adjudicator participants left some responses on the CAPS blank, the researcher chose the option in SPSS to replace those missing values with ‘replace with mean’ to compensate for those missing values. By doing this, SPSS substituted each missing item with the mean value of the other choral adjudicator participant responses for that item.

A Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett’s Test of Sphericity were used to test the assumptions of sampling adequacy (Kaiser, 1974). *Laerd Statistics* (2015a) stated, “Its value [for KMO] can range from 0 to 1, with values above 0.6 suggested as a minimum requirement for sampling adequacy, but values above 0.8 considered good and indicative of principal components analysis being useful” (p. 7). The overall KMO for all 53 items from choral adjudicator participant responses on the CAPS was .571 with Bartlett’s Test of Sphericity being significant ( $p = .000$ ). By using all 53 items on the CAPS the assumption for sampling adequacy was not met.

Use of all 53 items from the CAPS was determined not to be appropriate for a PCA. Review of the scree plot for all 53 items revealed the inflection point at the fourth factor (see Figure 2). *Laerd Statistics* (2015a) described the inflection point as representing “the point where the graph begins to level and subsequent components [factors] add little to the total variance” (p. 10). Based on these criteria the researcher determined it was appropriate to retain four factors. A PCA was calculated using forced factor extraction with four factors.

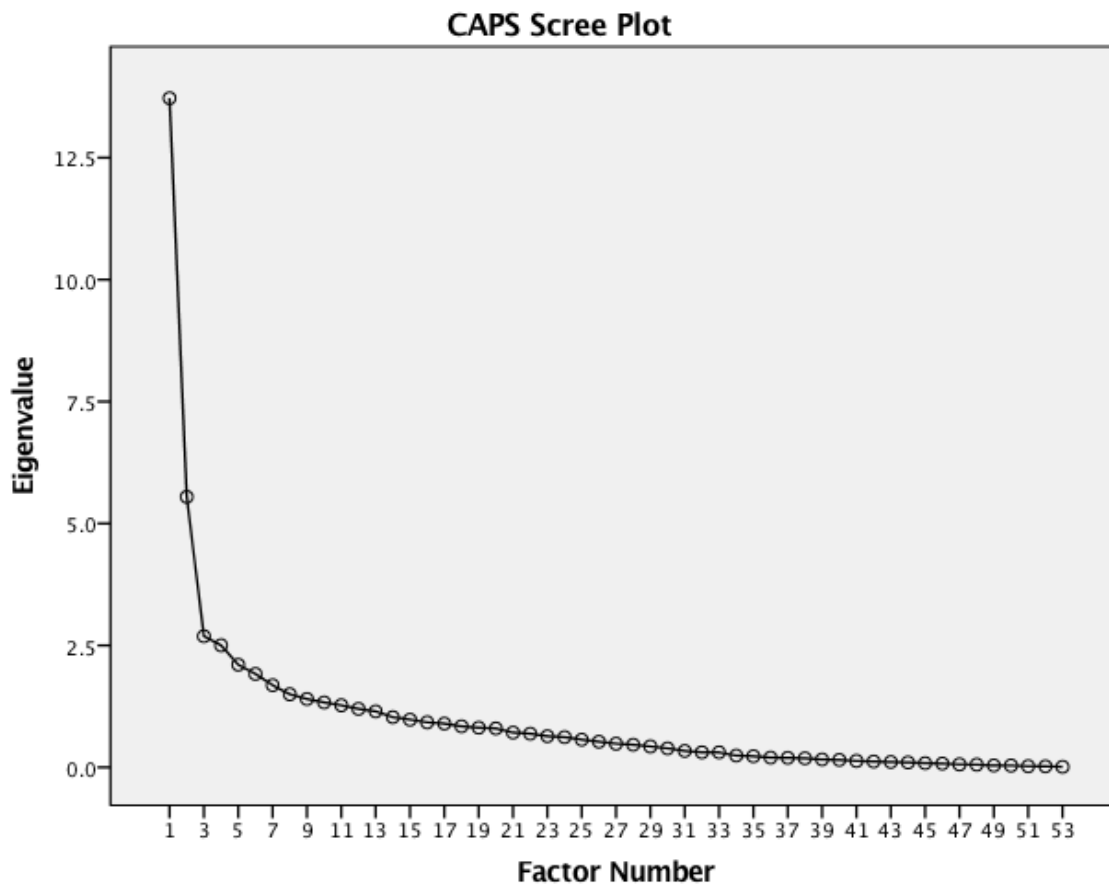


Figure 2. Scree Plot for *Choral Adjudicator Preference Scale* Responses.

Asmus (1989) suggested a participant-to-item ratio for factor analysis of 3:1. Therefore, the number of participants in the present study ( $N = 71$ ) required the number of items to be considered in the PCA be limited to 23. Community is the degree to which an item correlates with all other items, and it is often used to eliminate items with low values, as the goal of factor analysis is to explain the variance through the common factors (Child, 2006). Using community values, the top 23 items were selected for use in the PCA (see Table 7).

Table 7

## Communalities

Item Rank	Item Description	Initial	Extraction
1	Harmonic intonation	1	.697
2	Accuracy of rhythm	1	.658
3	Appearance	1	.654
4	Melodic intonation	1	.609
5	Ensemble size	1	.604
6	First impression	1	.598
7	How the ensemble walks off the stage	1	.597
8	Phrasing	1	.591
9	Ensemble's responsiveness to the conductor	1	.585
10	Vocal technique	1	.585
11	Performance space	1	.584
12	How the ensemble walks onto the stage	1	.569
13	Balance	1	.562
14	Accuracy of pitches	1	.548
15	Context of the performance event (purpose and location)	1	.548
16	Conducting gesture	1	.547
17	Musical style	1	.539
18	Time of day	1	.534
19	Breath support	1	.517
20	Age level	1	.510
21	Interpretation	1	.491
22	Tone quality	1	.485
23	Appropriateness of consonants	1	.481

Research Question 1 addressed factors that most influenced choral adjudicator participants' rating decisions. The 23 highest ranked items, based on communality value, were used to test the assumptions of a PCA. For the assumption of linearity *Laerd*

*Statistics* (2015a) states, “The level of correlation considered worthy of a variable’s inclusion is usually  $r \geq .3$ ” (p. 7). A review of the correlation matrix revealed that all items correlated with at least one other item greater than the recommended minimum correlation ( $r \geq 0.3$ ) (see Table 8). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .780, designated to be a ‘middling’ level according to Kaiser (1974). Bartlett’s Test of Sphericity was found to be significant ( $p = .000$ ). Further review of items’ individual KMO measures revealed the lowest item KMO to be at .437 and the highest at .749 (see Table 9). These results indicated that data generated from the 23 top-ranked items were appropriate to be analyzed with a PCA procedure.

A PCA was conducted using the 23 highest ranked items from the *Choral Adjudicator Preference Scale* to reduce the data into a smaller list of more discrete factors and to reveal possible latent patterns among items within those factors. A direct oblimin rotation method was selected for the analysis. Fabrigar, Wegener, MacCallum, and Strahan (1999) determined that oblique rotations provided an accurate means to describe simple latent patterns, as well as common themes, among correlated factors. They stated, “In contrast to orthogonal rotations, oblique rotations permit correlations among factors...Several oblique rotations procedures are commonly used and have been found to generally produce satisfactory solutions” (p. 281). The researchers also stated, “Finally, oblique solutions provide more information than orthogonal rotations” (p. 282). The goal of PCA and using an oblique rotation method, was to uncover a simple structure in which factors were defined by subsets via item loadings (Thurstone, 1947).

Table 8

Correlation Matrix: 23 Items

	C	F	G	K	P	Q	U	W	AA	AB	AC	AF	AH	AJ	AK	AL	AM	AO	AP	AQ	AR	AS	AW
C	1.00																						
F	<b>.35</b>	1.00																					
G	.12	-.10	1.00																				
K	.16	<b>.39</b>	-.10	1.00																			
P	.13	<b>.32</b>	.14	.29	1.00																		
Q	.18	<b>.39</b>	.03	<b>.45</b>	<b>.52</b>	1.00																	
U	.25	.12	<b>.33</b>	<b>.30</b>	<b>.39</b>	<b>.30</b>	1.00																
W	.17	.16	<b>.40</b>	.04	<b>.30</b>	.01	<b>.36</b>	1.00															
AA	.26	.09	<b>.38</b>	.14	<b>.31</b>	.29	<b>.65</b>	.24	1.00														
AB	.18	<b>.46</b>	-.02	<b>.63</b>	<b>.38</b>	<b>.49</b>	.22	.21	.17	1.00													
AC	.11	<b>.31</b>	-.10	<b>.72</b>	<b>.42</b>	<b>.58</b>	<b>.31</b>	.12	.23	<b>.72</b>	1.00												
AF	-.03	.20	<b>.43</b>	.15	.18	.13	.21	<b>.40</b>	.17	.21	.08	1.00											
AH	.27	<b>.47</b>	-.04	<b>.35</b>	<b>.46</b>	<b>.53</b>	.15	.23	.28	<b>.48</b>	<b>.43</b>	.22	1.00										
AJ	<b>.45</b>	.08	<b>.43</b>	.18	.25	.26	<b>.53</b>	<b>.39</b>	<b>.31</b>	.23	.20	.16	.17	1.00									
AK	<b>.38</b>	<b>.32</b>	.21	<b>.38</b>	<b>.47</b>	<b>.50</b>	<b>.42</b>	.22	<b>.51</b>	<b>.46</b>	<b>.41</b>	.22	<b>.31</b>	<b>.47</b>	1.00								
AL	.17	<b>.46</b>	-.13	<b>.37</b>	<b>.32</b>	<b>.42</b>	.05	.14	.05	<b>.61</b>	<b>.58</b>	.12	<b>.57</b>	.23	<b>.32</b>	1.00							
AM	<b>.32</b>	.10	.20	.14	.21	<b>.36</b>	<b>.53</b>	.21	<b>.75</b>	.19	<b>.30</b>	.16	.18	.26	<b>.55</b>	.14	1.00						
AO	.23	-.06	<b>.55</b>	-.05	.21	.03	<b>.45</b>	<b>.45</b>	<b>.30</b>	-.14	-.04	.24	.04	<b>.56</b>	.23	.00	<b>.33</b>	1.00					
AP	<b>.39</b>	.17	.20	.26	<b>.35</b>	.21	<b>.37</b>	.28	.26	.24	<b>.31</b>	.10	<b>.46</b>	<b>.46</b>	.19	.26	.18	.28	1.00				
AQ	-.04	<b>.16</b>	<b>.42</b>	-.04	.20	.02	.23	<b>.45</b>	.10	.03	.02	<b>.42</b>	.07	.23	.04	-.01	.12	<b>.43</b>	.07	1.00			
AR	.13	<b>.56</b>	-.02	<b>.64</b>	<b>.42</b>	<b>.42</b>	.26	.19	.13	<b>.69</b>	<b>.65</b>	.21	<b>.35</b>	.26	<b>.39</b>	<b>.55</b>	.17	.00	<b>.35</b>	.12	1.00		
AS	<b>.31</b>	.10	<b>.40</b>	.19	<b>.38</b>	<b>.35</b>	<b>.64</b>	<b>.38</b>	<b>.57</b>	.19	.19	.24	.18	<b>.51</b>	<b>.43</b>	.05	<b>.56</b>	<b>.40</b>	<b>.31</b>	<b>.32</b>	.24	1.00	
AW	.09	<b>.33</b>	.05	<b>.42</b>	<b>.43</b>	<b>.49</b>	.18	.23	.16	<b>.55</b>	<b>.56</b>	.20	<b>.51</b>	.28	<b>.42</b>	<b>.58</b>	.19	.13	.24	.23	<b>.39</b>	.16	1.00



Table 9

*Choral Adjudicator Preference Scale: Items' Individual Kaiser-Meyer-Olkin (KMO)*

Item Description	KMO Measure
How the ensemble walks off the stage	.749
How the ensemble walks on the stage	.724
Accuracy of rhythm	.721
Harmonic intonation	.719
Interpretation	.669
Appearance	.661
Conducting gesture	.656
Ensemble size	.653
First impression	.650
Performance space	.649
Vocal technique	.625
Melodic intonation	.624
Time of day	.611
Ensemble's responsiveness to the conductor	.598
Musical style	.589
Balance	.577
Context of the performance event (purpose and location)	.577
Accuracy of pitches	.566
Age level	.562
Breath support	.541
Phrasing	.535
Tone quality	.450
Appropriateness of consonants	.437

Researchers described the complex nature of factor rotation methods. Fabrigar et al. (1999) stated,

For any given solution with two or more factors (or principal components), there exists an infinite number of alternative orientations of the factors in multidimensional space that will explain the data equally well . . . Therefore, a researcher must select a single solution from among the infinite number of equally fitting solution. (p. 281)

Osborne and Costello (2009) suggested possible procedures for exploring PCA, including:

. . . run the data four times, setting the number of factors extracted at four, five, six, and seven. After rotation . . . compare the item loading tables; the one with the “cleanest” factor structure—item loadings above .30, no or few item crossloadings, no factors with fewer than three items—has the best fit to the data. (p. 135)

Considering the descriptions and suggestions of Fabrigar et al. (1999) and Osborne and Costello (2009), all possible orientations between two and five-factor rotations were explored using principal component analysis with direct oblimin rotations and delta parameter adjustments. The preliminary analyses resulted in a total of 68 computations of the direct oblimin rotation method. The direct oblimin rotation method ( $\delta = 0$ ), while extracting four factors, resulted in the cleanest factor structure (see Table 10). The principal component analysis using the highest ranked 23 items from the CAPS converged in 12 iterations and suggested four factors each with eigenvalues greater than one.

Table 10

## Loadings and Variance Contributions for Four-Factor Structure Matrix

Factor Name/Items	Factor Number			
	1	2	3	4
<b>1. The Ensemble's Performance</b>				
Harmonic intonation	<b>.845</b>	.051	.004	-.065
Accuracy of rhythm	<b>.803</b>	.209	-.107	-.106
Melodic intonation	<b>.776</b>	.006	.082	.006
Vocal technique	<b>.776</b>	-.191	-.018	.160
Accuracy of pitches	<b>.725</b>	.142	-.117	-.072
Breath support	<b>.693</b>	-.015	.210	-.007
Tone quality	<b>.647</b>	-.159	.050	.147
Phrasing	<b>.647</b>	-.069	.027	.266
Balance	<b>.619</b>	.363	-.114	-.056
Appropriateness of consonants	<b>.485</b>	.224	.198	.055
Variance contributed by Factor 1 = 31.937%				
<b>2. Visual Aspects</b>				
How the ensemble walks onto the stage	.021	<b>.867</b>	-.071	-.023
How the ensemble walks off the stage	-.029	<b>.866</b>	-.006	.022
Appearance	.034	<b>.683</b>	.164	.157
First impression	-.002	<b>.655</b>	.231	.150
Ensemble's responsiveness to the conductor	.374	<b>.572</b>	-.034	.083
Variance contributed by Factor 2 = 15.770%				
<b>3. Extra-musical Aspects</b>				
Ensemble size	.033	-.058	<b>.832</b>	-.101
Age level	.217	.027	<b>.716</b>	-.239
Context of the performance event (purpose and location)	.093	-.012	<b>.663</b>	.230
Time of day	-.253	.255	<b>.636</b>	.122
Performance space	-.250	.217	<b>.523</b>	.397
Variance contributed by Factor 3 = 7.922%				
<b>4. The Conductor's Contributions</b>				
Interpretation	.041	.134	-.237	<b>.783</b>
Musical style	.229	-.025	.039	<b>.689</b>
Conducting gesture	.045	.221	.231	<b>.612</b>
Variance contributed by Factor 4 = 5.866%				
<b>Total variance contributed by all 4 Factors = 61.495%</b>				

With four discrete factors identified, the combination of items loaded on each factor was analyzed to determine what the items had in common for that factor. The ten items correlated on Factor One related to the ensemble's performance specifically and explained 31.937% of the total variance in choral adjudicator participant responses. The five items that correlated on Factor Two related to visual aspects and explained 15.770% of the total variance in choral adjudicator participant responses. The four items that correlated on Factor Three related to extra-musical aspects and explained 7.922% of the total variance in choral adjudicator participant responses. The three items that correlated on Factor Four related to the conductor's contributions specifically and explained 5.866% of the total variance in choral adjudicator participant responses. When summed together, all factors explained 61.495% of the total variance in choral adjudicator participant responses (see Table 11).

Table 11

## Total Variance Explained: Four Factors

Primary Factor Number and Name	Eigenvalues	Percentage of Variance	Cumulative Percentage
One: The Ensemble's Performance	7.345	31.937	31.937
Two: Visual Aspects	3.627	15.770	47.706
Three: Extra-musical Aspects	1.822	7.922	55.629
Four: The Conductor's Contributions	1.349	5.866	61.495

## **Research Question 2: Years of Adjudication Experience**

Research question two was: “Do differences exist among adjudicators’ influences on rating decisions by years of adjudication experience?” Data from the *Choral Adjudicator Preference Scale* (CAPS) and the four factors revealed in research question one were used to answer Research Question 2. A two-way repeated measures analysis of variance (ANOVA) was calculated to determine if differences existed among choral adjudicators influences on ratings of choral performance quality by the four factors by years of adjudication experience. IBM’s SPSS software was employed to calculate the two-way repeated measures ANOVA. Microsoft Excel 2011 was used to aid in organizing the data calculated by SPSS. *Laerd Statistics* (2015b) was also used as a guide for aiding the process and procedure for the analysis.

Descriptive statistics (i.e., means and standard deviations) were calculated for participants’ groupings by Years of Adjudication Experience (YAE) and organized under each primary factor (see Table 12). Although there was a total of 71 participants who completed the *Choral Adjudicator Preference Scale* (CAPS) one choral adjudicator participant did not answer the item “How many years have you been adjudicating choral performance, including the current school year?” in the demographic section. That participant’s responses were not included in the two-way repeated measures analysis, which resulted in a total of 70 participants’ responses being used for the analysis.

Two independent variables, a within-subjects variable and a between-subjects variable, were examined in the two-way repeated measures ANOVA. Participants’

responses to the items on the *Choral Adjudicator Preference Scale* (CAPS) provided the dependent measure for both independent variables.

Table 12

## Descriptive Statistics: Factors and Years of Adjudication Experience

Factor Description	Years of Adjudication Experience			
		<i>N</i>	<i>M</i>	<i>SD</i>
The Ensemble's Performance	0 to 5	10	4.35	.580
	6 to 10	17	4.50	.498
	11 to 15	8	4.46	.644
	16 to 20	17	4.57	.380
	21 to 25	5	4.50	.474
	26 to 30	3	4.70	.265
	30 or more	10	4.31	.415
	Total	70	4.47	.475
Visual Aspects	0 to 5	10	2.30	.701
	6 to 10	17	2.71	.834
	11 to 15	8	2.95	.880
	16 to 20	17	2.83	.925
	21 to 25	5	2.84	.932
	26 to 30	3	3.00	.693
	30 or more	10	2.80	.706
	Total	70	2.74	.819
Extra-musical Aspects	0 to 5	10	1.76	.460
	6 to 10	17	2.10	.732
	11 to 15	8	2.07	1.047
	16 to 20	17	1.50	.656
	21 to 25	5	1.84	.841
	26 to 30	3	1.86	.231
	30 or more	10	1.42	.494
	Total	70	1.78	.715
The Conductor's Contributions	0 to 5	10	3.03	.744
	6 to 10	17	3.17	.667
	11 to 15	8	3.16	.643
	16 to 20	17	3.39	.892
	21 to 25	5	3.33	.625
	26 to 30	3	3.22	.387
	30 or more	10	3.30	.674
	Total	70	3.23	.709

The within-subjects variable for the two-way repeated measures ANOVA (Factors) included four levels (i.e., The Ensemble’s Performance, Visual Aspects, Extra-musical Aspects, and The Conductor’s Contributions). These factors were derived from the principal component analysis used to answer research question one. The between-subjects variable for the two-way repeated measures ANOVA was participants’ Years of Adjudication Experience (YAE) and included seven levels. Using the CAPS demographic section item “How many years have you been adjudicating choral performance, including the current school year?” participants were grouped by years of choral adjudication experience (see Table 13).

Table 13

## Participant Groupings by Years of Adjudication Experience

Participant Group (YAE) <sup>a</sup>	Number	Percentage
0-5	10	14.28%
6-10	17	24.28%
11-15	8	11.42%
16-20	17	24.28%
21-25	5	7.14%
26-30	3	4.28%
30 or more	10	14.28%
Total	70	100.00%

<sup>a</sup>YAE is an abbreviation for Years of Adjudication Experience

Ten participants had zero to five years of choral adjudication experience, which represented 14% of the total sample of participants. Seventeen participants had 6–10 years of choral adjudication experience, which represented 24% of the total sample of participants. Eight participants had eleven to fifteen years of choral adjudication

experience, which represented 11% of the total sample of participants. Seventeen participants had 16–20 years of choral adjudication experience, which represented 24% of the total sample of participants. Five participants had 21–25 years of choral adjudication experience, which represented 7% of the total sample of participants. Three participants had 23–30 years of choral adjudication experience, which represented 4% of the total sample of participants. Ten participants had 30 or more years of choral adjudication experience, which represented 14% of the total sample of participants.

Using individual participant CAPS responses for items loaded on each primary factor derived from research question one, a mean-score was calculated for each participant on each primary factor. Primary factor mean-scores were used to describe how each participant's rating decision of a choral performance was influenced by that factor (see Table 12). Mean scores for choral adjudicator participant groups for Factor One (The Ensemble's Performance) ranged from 4.31 to 4.70 with a total participant mean-score of 4.48 and a standard deviation of 0.475. Choral adjudicator participants reported that 'The Ensemble's Performance' had a very strong to extremely strong influence on their rating decision of choral performance. Mean scores for choral adjudicator participant groups for Factor Two (Visual Aspects) ranged from 2.30 to 3.00 with a total participant mean-score of 2.75 and a standard deviation of .819. Choral adjudicator participants reported that 'Visual Aspects' had a somewhat strong to strong influence on their rating decision of choral performance. Mean scores for choral adjudicator participant groups for Factor Three (Extra-musical Aspects) ranged from 1.420 to 2.106 with a total participant mean-score of 1.780 and a standard deviation of



.715. Choral adjudicator participants reported that ‘Extra-musical Aspects’ had a not strong to somewhat strong influence on their rating decision of choral performance. Mean scores for choral adjudicator participant groups for Factor Four (The Conductor’s Contributions) ranged from 3.03 to 3.39 with a total participant mean-score of 3.24 and a standard deviation of .709. Choral adjudicator participants reported that ‘The Conductor’s Contributions’ had a strong influence on their rating decision of choral performance.

The purpose of the second research question was to investigate a possible interaction effect among factors influencing participants’ ratings of choral performance by participants’ years of adjudication experience. A two-way repeated measures ANOVA was performed, with the Factors serving as the within-subjects variable and participants’ Years of Adjudication Experience (YAE) serving as the between-subjects variable. Mauchly’s Test of Sphericity, used to test the assumption that the variances of the differences between related groups are equal, was not violated ( $p = .788$ ). A significant main effect was found for Factors,  $F(3, 189) = 216.581, p = .000, \eta^2 = .775$ . However, no significant main effect was found for YEA and no significant interaction effect was found among Factors by YEA (see Table 14).

Pairwise comparisons revealed significant differences among within-subjects factors, including Factor One (The Ensemble’s Performance) and Factor Two (Visual Aspects) ( $p = .000$ ), Factor One (The Ensemble’s Performance) and Factor Three (Extra-musical Aspects) ( $p = .000$ ), Factor One (The Ensemble’s Performance) and Factor Four (The Conductor’s Contributions) ( $p = .000$ ), Factor Two (Visual Aspects) and Factor

Three (Extra-musical Aspects) ( $p = .000$ ), and Factor Two (Visual Aspects) and Factor Four (The Conductor's Contributions) ( $p = .000$ ), and Factor Three (Extra-musical Aspects) and Factor Four (The Conductor's Contributions) ( $p = .000$ ) (see Table 15).

Table 14

Two-Way Repeated Measures ANOVA: Factors by Years of Adjudication Experience

Source <sup>a</sup>	SS	df	MS	F	Sig	Partial Eta <sup>2</sup>	Noncent. Parameter	Observed Power
Factors	187.799	3	62.600	216.581	.000	.775	649.742	1.000
YAE	2.990	6	.498	.464	.832	.042	2.787	.177
Factors*YAE	6.648	18	.369	1.278	.206	.108	22.999	.829
Error	54.628	189	.289					

<sup>a</sup>YAE is an abbreviation for Years of Adjudication Experience

Table 15

Pairwise Comparisons: Within-Subjects Factors

(a) Factors	(b) Factors	Mean Difference (a-b)	Std. Error	Sig. <sup>b</sup>
1	2	1.70*	.111	.000
	3	2.69*	.106	.000
	4	1.25*	.096	.000
2	1	-1.70*	.111	.000
	3	.98*	.111	.000
	4	-.45*	.113	.000
3	1	-2.69*	.106	.000
	2	-.98*	.111	.000
	4	-1.43*	.106	.000
4	1	-1.25*	.096	.000
	2	.45*	.113	.000
	3	1.43*	.106	.000

*Note.* \*The mean difference is significance at the .05 level

<sup>b</sup>Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments)

### **Research Question 3: Academic Training**

Research question three was: “Do differences exist in adjudicators’ influences on rating decisions by academic training?” Data from the *Choral Adjudicator Preference Scale* (CAPS) and the four factors revealed in research question one were used to answer research question three. A two-way repeated measures analysis of variance (ANOVA) was calculated to determine if differences existed among choral adjudicators’ influences on ratings of choral performance quality by the four factors by participants’ academic training. IBM’s SPSS software was employed to calculate the two-way repeated measures ANOVA. Microsoft Excel 2011 was used to aid in organizing the data calculated by SPSS. *Laerd Statistics* (2015b) was also used as a guide for aiding the process and procedure for the analysis.

Descriptive statistics (i.e., means and standard deviations) were calculated for participants’ grouping by academic training and organized under each factor (see Table 16). Although there was a total of 71 participants who completed the *Choral Adjudicator Preference Scale* (CAPS) one choral adjudicator participant did not answer the item “Please list academic degrees earned as well as their concentration (e.g., Bachelor Degree, Music Education; Master Degree, Vocal Performance; Doctorate in the Musical Arts, Conducting)” in the demographic section. That participant’s responses were not included in the two-way repeated measures ANOVA, which resulted in a total of 70 participants’ responses being used for the analysis.

Table 16

## Descriptive Statistics: Factors and Academic Training

Factor Description	Academic Training	<i>N</i>	<i>M</i>	<i>SD</i>
The Ensemble's Performance	1	37	4.54	.412
	2	24	4.43	.510
	3	9	4.17	.547
	Total	70	4.46	.474
Visual Aspects	1	37	2.74	.832
	2	24	2.49	.787
	3	9	3.13	.748
	Total	70	2.70	.821
Extra-musical Aspects	1	37	1.80	.650
	2	24	1.58	.621
	3	9	2.11	1.087
	Total	70	1.76	.717
The Conductor's Contributions	1	37	3.26	.654
	2	24	3.07	.804
	3	9	3.29	.633
	Total	70	3.20	.703

Two independent variables, a within-subjects variable and a between-subjects variable, were examined in the two-way repeated measures ANOVA. Participants' responses to the items on the *Choral Adjudicator Preference Scale* (CAPS) provided the dependent measure for both independent variables.

The within-subjects variable for the two-way repeated measures ANOVA (Factors) included four levels (i.e., The Ensemble's Performance, Visual Aspects, Extra-musical Aspects, and The Conductor's Contributions). These factors were derived from the principal component analysis used to answer research question one. The between-subjects variable for the two-way repeated measures ANOVA was choral adjudicator participants' academic training and included three levels. Using the CAPS demographic

section item “Please list academic degrees earned as well as their concentration (e.g., Bachelor Degree, Music Education; Master Degree, Vocal Performance; Doctorate in the Musical Arts, Conducting)” the researcher grouped participants into three groups: Group One, music education concentration; Group Two, music performance concentration; and Group Three, hybrid classification—music education and performance (see Table 17). Group One consisted of thirty-seven participants who had music education degrees, which represented 53% of the total sample of participants. Group Two consisted of twenty-four participants who had music performance degrees, which represented 34% of the total sample of participants. Group Three consisted of nine participants who had a ‘hybrid’ classification, where they had both music education and performance degrees, which represented 13% of the total sample of participants.

Table 17

## Participant Groups by Academic Training

Participant Group (ACT) <sup>a</sup>	Number	Percentage <sup>b</sup>
(1) Education	37	53%
(2) Performance	24	34%
(3) Hybrid—Education and Performance	9	13%
Total	70	100%

<sup>a</sup>ACT is an abbreviation for Academic Training

<sup>b</sup>Percentage is rounded to the nearest whole number

Using individual participant CAPS responses for items loaded on each primary factor derived from research question one, a mean score was calculated for each participant on each primary factor. Primary factor mean scores were used to describe

how each participant's rating decision of a choral performance was influenced by that factor (see Table 15). Mean scores for choral adjudicator academic training groups for The Ensemble's Performance ranged from 4.18 to 4.55 with a total participant mean-score of 4.46 and a standard deviation 0.474. Choral adjudicator participants reported that 'The Ensemble's Performance' had a very strong influence on their rating decision of choral performance. Mean scores for choral adjudicator academic training groups for 'Visual Aspects' ranged from 2.49 to 3.13 with a total participant mean score of 2.71 and a standard deviation of .821. Choral adjudicator participants reported that 'Visual Aspects' had a somewhat strong to strong influence on their rating decision of choral performance. Mean scores for choral adjudicator academic training groups for 'Extra-musical Aspects' ranged from 1.58 to 2.11 with a total participant mean score of 1.77 and a standard deviation of .717. Choral adjudicator participants reported that 'Extra-musical Aspects' had a not strong to somewhat strong influence on their rating decision of choral performance. Mean scores for choral adjudicator academic training groups for 'The Conductor's Contributions' ranged from 3.07 to 3.30 with a total participant mean score of 3.20 and a standard deviation of .703. Choral adjudicator participants reported that 'The Conductor's Contributions' had a strong influence on their rating decision of choral performance.

The purpose of the third research question was to investigate a possible interaction effect among factors influencing participants' ratings of choral performance by participants' academic training. A two-way repeated measures ANOVA was performed, with the Factors serving as the within-subjects variable and participants'

Academic Training (ACT) serving as the between-subjects variable. Mauchly's Test of Sphericity, used to test the assumption that the variances of the differences between related groups are equal, was not violated ( $p = .359$ ). A significant main effect was found for the Factors,  $F(3, 201) = 195.326, p = .000, \eta^2 = .745$ . However, no significant main effect was found for ACT and no significant interaction effect was found among Factors by ACT (see Table 18).

Table 18

Two-Way Repeated Measures ANOVA: Factors by Academic Training

Source <sup>a</sup>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>Sig</i>	Partial Eta <sup>2</sup>	Noncent. Parameter	Observed Power
Factors	168.377	3	56.126	195.326	.000	.745	585.979	1.000
ACT	3.075	2	1.537	1.529	.224	.044	3.059	.314
Factors*ACT	3.297	6	.549	1.912	.080	.054	11.473	.699
Error	57.756	201	.287					

<sup>a</sup>ACT is an abbreviation for Academic Training.

Pairwise comparisons revealed significant differences among within-subjects factors, including Factor One (The Ensemble's Performance) and Factor Two (Visual Aspects) ( $p = .000$ ), Factor One (The Ensemble's Performance) and Factor Three (Extra-musical Aspects) ( $p = .000$ ), Factor One (The Ensemble's Performance) and Factor Four (The Conductor's Contributions) ( $p = .000$ ), Factor Two (Visual Aspects) and Factor Three (Extra-musical Aspects) ( $p = .000$ ), and Factor Two (Visual Aspects) and Factor Four (The Conductor's Contributions) ( $p = .000$ ), and Factor Three (Extra-musical Aspects) and Factor Four (The Conductor's Contributions) ( $p = .000$ ) (see Table 19).

Table 19

## Pairwise Comparisons: Within-Subjects Factors

(a) Factors	(b) Factors	Mean Difference (a-b)	Std. Error	Sig. <sup>b</sup>
1	2	1.59*	.103	.000
	3	2.55*	.104	.000
	4	1.17*	.092	.000
2	1	-1.59*	.103	.000
	3	.95*	.118	.000
	4	-.41*	.110	.000
3	1	-2.55*	.104	.000
	2	-.95*	.118	.000
	4	-1.37*	.114	.000
4	1	-1.17*	.092	.000
	2	.41*	.110	.000
	3	1.37*	.114	.000

*Note.* \*The mean difference is significance at the .05 level.

<sup>b</sup>Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

## Summary of Results

### Participants

Participants for the present study included individuals who serve as choral performance adjudicators within the Southern Division of the National Association for Music Education (NAfME) and the American Choral Directors Association (ACDA) ( $N = 71$ ). The Southern Division of NAfME and ACDA include the states of Alabama, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and West Virginia, and the Commonwealths of Kentucky and Virginia (National Association for Music Education, December 2015).



## **Methodology**

A Modified Delphi method was employed to determine salient influences on choral adjudicators' ratings of choral performance quality. That information was used to create an online survey for participants to complete. A random sample of choral adjudicators from the North Carolina Music Educators Association (NCMEA) was asked to respond to an open-ended questionnaire regarding influences on their rating decision of choral performance quality. Adjudicators' responses were analyzed and used to generate a list of the most prominent items, which was vetted by experts from professional music organizations within the Southern Divisions of NAfMEA and ACDA. The experts' feedback was used to create the *Choral Adjudicator Preference Scale (CAPS)* in the form of an online survey to collect data from participants (see Appendix E).

### **Research Question 1: Factors Influencing Choral Adjudicators' Rating Decision**

A principal component analysis was calculated using the 23 top-ranked items from a 53-question survey that measured choral adjudicator participants' self-reported influences on their rating decision of a choral ensemble's performance. The appropriateness of a principal component analysis was assessed prior to the analysis. Inspection of the correlation matrix showed that each item correlated with at least one other item with a correlation coefficient greater than .3. The overall Kaiser-Meyer-Olkin (KMO) measure was 0.780 indicating a classification of 'middling' according to Kaiser (1974). Bartlett's Test of Sphericity was statistically significant ( $p < .000$ ).

The four-factor solution explained 61.495% of the total variance. A direct oblimin oblique rotation method was employed to aid interpretability. The rotated

solution revealed a 'simple structure' (Thurstone, 1947). Interpretation of the data revealed four factors that influenced choral adjudicators ratings of choral performance: The Ensemble's Performance on Factor One, Visual Aspects on Factor Two, Extra-musical Aspects on Factor Three, and The Conductor's Contributions on Factor Four.

### **Research Question 2: Years of Adjudication Experience**

A two-way repeated measures ANOVA was calculated to determine the effect of choral adjudicators' years of adjudication experience on their rating decision of choral performance factors. The analysis revealed a significant main effect for Factors among participants ( $p = .000$ ), however; there was no significant main effect for Years of Adjudication Experience among participants and no significant interaction effect among Factors by participants' Years Adjudication of Experience.

### **Research Question 3: Academic Training**

A two-way repeated measures ANOVA was calculated to determine the effect of choral adjudicators' academic training on their rating decision of choral performance factors. The analysis revealed a significant main effect for Factors among participants ( $p = .000$ ), however; there was no significant main effect for participant's Academic Training and no significant interaction effect among Factors by participants' Academic Training.

## **CHAPTER V**

### **CONCLUSIONS**

#### **Summary**

The issue that motivated the present research study was concerns regarding adjudicator bias in the context of choral ensemble performance adjudication. Researchers have suggested the level of agreement among music performance adjudicators has been typically less than desirable (Bergee & Platt, 2003; Latimer, 2007) and that the cause might be rooted in adjudicators' bias (McPherson & Thompson, 1998; Payne, 1997). The present research study aimed to (a) illuminate factors that influenced adjudicators' ratings of ensemble performance quality and to (b) investigate possible bias by years of adjudication experience and academic training.

Historically, music contests and festivals have been used to stimulate teaching excellence in the school system (Rohrer, 2002). Since the early 20<sup>th</sup> century, music contests and festivals have been an expectation for music educators and students of music ensembles across the country (Freer, 2011; Guegold, 1989; Rohrer, 2002). As stakeholders of music programs continued to use the music contest and festival as a measure of teaching and learning effectiveness (Corbin, 1995; Forbes, 1994; Freer, 2011) concerns regarding adjudicator bias grew (Cassidy & Sims, 1991; Cheek, 2007; Corbin, 1995; Elliot, 1995/1996; Fox, 1990; McPherson & Thompson, 1998; Radocy, 1976; Robinson, 1990). These concerns regarding adjudicator bias provided the motivation to

investigate influences on adjudicators' rating decisions of choral performance quality. The intent of the present study was to investigate the source of the performance rating (i.e., the adjudicators themselves) rather than the measures used to evaluate choral performance or the environment of the performance setting.

McPherson and Thompson's (1998) Process Model of Assessing Music Performance (PMAMP) provided the framework for inquiry of music performance assessment. The findings of the present study are positioned within the PMAMP's fourth factor, Evaluator Characteristics. The findings of the present study contribute to the greater body of literature regarding choral music performance assessment and the evaluation of music performance assessment.

A modified Delphi method was used to create the *Choral Adjudicator Preference Scale* (CAPS). Data collected via the CAPS were then used to answer research questions one through three. For question one, a Principal Component Analysis (PCA) was calculated to determine whether groups of items that choral adjudicators identified as influencing their rating decisions would coalesce under a smaller number of discrete factors. To answer questions two and three, a two-way repeated measures analysis of variance (ANOVA) was calculated using demographic data from the CAPS and the factors derived from the PCA.

Results from the present research study should be interpreted with a degree of caution. The researcher received a total of 287 email contacts for choral adjudicators in the Southeast. Of the 287 emails 65 did not work due to incorrect email addresses, email addresses that were no longer used, or limitations among email servers' acceptance of

emails from UNCG's Qualtrics. As a result, only 222 email contacts were successfully sent. Although the researcher sent eight invitations to choral adjudicators in the Southeast, there was no opportunity to determine exactly how many possible participants actually received the invitation. Ninety-four participants accepted the invitation to participate. Of those 94 only 71 participants completed the CAPS measure, a 75% response rate from those who accepted the invitation to participate. More participants might reveal more factors due to the need to maintain a proper (3:1) subject-to-variable ratio. More participants might reveal differing results for the PCA and consequently for the two two-way repeated measures ANOVAs. Nevertheless, results from the present study are informative and provide insight into influences on choral adjudicators' rating decisions.

Results provided evidence of four factors of influence on choral adjudicators' rating decisions of choral performance quality. There were significant differences among the factors that influence adjudicators' rating decisions of ensemble performance quality; however, interactions with adjudicators' years of adjudication experience and academic training were not found. Essentially, differences among the factors were not due to adjudicators' years of adjudication experience or academic training (degree focus). The present study provides evidence that adjudicators' adjudication experience and academic training might not interact with the factors that influence choral adjudicators' performance rating. More research in the domain of music performance assessment, specifically regarding characteristics of the evaluator, is recommended to identify reasons

for differences of influence among adjudicators. The sequence of research questions one through three will provide the structure for the remainder of this chapter.

### **Research Question 1: Factors Influencing Choral Adjudicators' Rating Decision**

Research Question 1 was: "What factors influence adjudicators' decisions when adjudicating choirs?" This research question was in direct response to McPherson and Thompson's (1998) call for further research when they stated, "Assessing musical performance is common across many types of music education practices, yet research clarifying the range of factors which impact on a judge's assessment is relatively scarce" (p. 12). Results from the present study suggested factors that influence adjudicators' ratings of choral performance quality. These factors supported results from previous research regarding influences on peoples' perceptions of music quality.

Four factors that influenced adjudicators' ratings of performance quality were (a) the ensemble's performance, (b) visual aspects, (c) extra-musical aspects, and (d) the conductor's contributions. Each factor was comprised of items that further explained specific aspects of influence. Each factor, and subsequent items within each factor, will be discussed.

#### **Factor One—The Ensemble's Performance**

Not surprisingly, the most important influence on an adjudicator's rating of a choral performance was the choral ensemble's performance itself. This factor accounted for most of the variance (31.93%) among the factors that influenced adjudicators' ratings. This finding might seem axiomatic; however, it is important because it provides evidence

that the ensemble's performance is at the core of music performance assessment (McPherson & Thompson, 1998).

Items loaded on Factor One (The Ensemble's Performance) seemed to represent categories one might find on a standardized choral-music performance rubric or rating form: (a) singing in tune (harmonic and melodic intonation), (b) tone and vocal production (vocal technique, breath support, and tone quality), (c) performing with precision (accurate pitches and rhythms, appropriate consonants), and (d) musicianship (balance and phrasing). All of these aspects work in tandem to represent The Ensemble's Performance.

Item loadings, represented by a correlation coefficient, provide valuable information regarding the magnitude [strength] of an item on the factor. Items loaded on Factor One (The Ensemble's Performance) were (listed in order): harmonic intonation ( $r = .845$ ), accuracy of rhythm ( $r = .803$ ), melodic intonation ( $r = .776$ ), vocal technique ( $r = .776$ ), accuracy of pitches ( $r = .725$ ), breath support ( $r = .693$ ), tone quality ( $r = .647$ ), phrasing ( $r = .647$ ), balance ( $r = .619$ ), and appropriateness of consonants ( $r = .485$ ).

The aspects mentioned previously are similar to Cooksey's (1977) findings. Cooksey's seminal research determined seven factors of choral performance: diction, precision, dynamics, tone control, tempo, balance and blend, and interpretation and musical effect. Cooksey's findings became the framework for many choral performance rubrics and ratings forms. Perhaps his findings influenced how adjudicators listened to choirs over time and contributed to how participants in the present study responded to items on the CAPS. Another possibility might be that adjudicators have always valued

these aspects of performance and that the current study provided further support that these aspects of choral performance continue to be valuable factors of influence.

Jones's (1986) study sought to develop a rating scale for high school vocal solo performance. Although Jones's work did not focus on the ensemble characteristics, the research was in the vocal performance domain. Jones' findings were similar to those of the present study. Jones's study produced a 32-item scale based on five factors: interpretation/musical effect, tone/musicianship, technique, suitability/ensemble, and diction. Factor One (The Ensemble's Performance) of the present study was similar to Jones's findings; items loaded on The Ensemble's Performance included several factors revealed in Jones's study.

### **Factor Two—Visual Aspects**

The second influence on adjudicators' rating of choral performance was visual aspects of the ensemble. This factor accounted for 15.77% of the variance among the factors that influenced adjudicators' ratings. This finding is supported by prior research in the domain of music perception and cognition regarding visual information's influence on peoples' perception of music quality (Howard, 2012; Juchniewicz, 2008; Mitchell & MacDonald, 2016; Platz & Kopiez, 2013; Tsay, 2013).

Items loaded, with correlation coefficients, on Factor Two (Visual Aspects) were ordered as follows: How the ensemble walks onto the stage ( $r = .867$ ), how the ensemble walks off the stage ( $r = .866$ ), appearance ( $r = .683$ ), first impression ( $r = .655$ ), and ensemble's responsiveness to the conductor ( $r = .572$ ).



These results in the present study are supported by prior research in that visual aspects of a choral ensemble might influence peoples' perceptions of music performance quality. Mitchell and MacDonald (2016) argued that visual information complimented audio information and transmitted music performers' intentions and quality to an audience. Tsay (2013) concluded that people depended primarily on visual information when making judgments regarding music performance quality. Platz and Kopiez (2013) determined that 'first impression' during the stage entrance had an influence on adjudicators' evaluations of music performance. Juchniewicz (2008) discovered that an ensemble's appearance, via movement during the performance, increased ratings of phrasing, dynamics, sense of rubato, and overall musical performance. Howard (2012) concluded that performers' overall appearance, including choice of clothing and how they conducted themselves on stage, affected performance quality ratings.

The two highest correlated items on Factor Two (Visual Aspects) were 'How the ensemble walks onto the stage' ( $r = .867$ ) and 'how the ensemble walks off the stage' ( $r = .866$ ). Both items had strong correlation coefficients and were close to one another, separated by only .001. Considering Mitchell and MacDonald's (2016) and Howard's (2012) findings, perhaps these items reveal that adjudicators are influenced by what the ensemble exudes, regarding the intent and quality of their performance, observed via stage deportment, when the members enter and exit the stage. Choral directors might find this finding to be informative and consequently may encourage ensemble members to carry themselves in such a manner that exhibits a sense of dignity, pride, and confidence.

An interesting discovery of the present study was the item ‘ensemble’s responsiveness to the conductor’ and its influence on Factor Two (Visual Aspects). This finding might be novel, as it was not found in prior research. The visual aspect of the choral ensemble, being responsive to the conductor, influences choral adjudicators ratings decisions. This determination makes sense when considering the findings of other researchers, those of Mitchel and MacDonald (2016) specifically. The visual aspect of the choral ensemble being sensitive and responsive to the conductor shows performers’ intent to being musical in their performance. The item acknowledges the importance of ensemble members’ active attentiveness and collaboration with the conductor. The item also acknowledges a connection between the ensemble and the conductor. Choral adjudicators wish to see an authentic connection among ensemble members and the conductor, one that might enable a high quality musical performance. Choral conductors and ensembles might find this information to be useful in preparing for music contests, festivals, and state MPA events.

### **Factor Three—Extra-musical Aspects**

The third influence on adjudicators’ rating of choral performance was extra-musical aspects. This factor accounted for 7.92% of the variance among the factors that influenced adjudicators’ ratings. This finding is supported by prior research that suggested non-musical, or extra-musical aspects of musical performance influenced peoples’ perceptions and ratings of music performance quality (Bergee, 2007; Bergee & Platt, 2003; Bergee & Westfall, 2005; James et al., 1984; McPherson & Thompson, 1998).

Items loaded, with correlation coefficients, on Factor Three (Extra-musical Aspects) were (in order): Ensemble size ( $r = .832$ ), age level ( $r = .716$ ), context of the performance (purpose and location) ( $r = .663$ ), time of day ( $r = .636$ ), and performance space ( $r = .523$ ). Ensemble size and age level had strong correlations with Factor Three. Context of the performance event, time of day, and the performance space had moderate correlations with Factor Three.

Ensemble size and age level are characteristics of the performers. Ensemble size is supported by McPherson and Thompson (1998), who suggested that the size of an ensemble had an influence in music performance and assessment. Cassidy and Sims (1991) supported age level as an influence. In their study, they reported a three-way interaction among label condition, presentation mode, and age. The amount of available research in the domain of music performance assessment regarding the influence of ensemble size and age level seems to be lacking. The finding of the present study provides further evidence that these two items are associated strongly with choral adjudicators' ratings of choral performance.

Perhaps adjudicators take into account ensemble size and age level regarding the appropriateness of the music and the performance when determining performance ratings. Cassidy and Sims (1991) suggested adjudicators' bias and prior knowledge regarding performers influenced adjudicator ratings. It was interesting to see that Factor Three supported the influence of some demographic information regarding the choir, ensemble size and age level of the singers. Although these items are not specifically 'prior knowledge,' as they can be noticed by simply observing the choir, they do provide

contextual information about the ensemble that can inform, or prime (Mitchell & MacDonald, 20016) adjudicators' expectations (Brown & Novak, 2007).

Prior researchers investigated and tested how the context of a performance, time of day, and performance space affect peoples' perceptions and ratings of performance (Bergee & Platt, 2003; McPherson, 1995; McPherson & Thompson, 1998). Participants in the present study indicated that the context of a performance influenced their ratings of performance quality. McPherson's (1995) study validated the presence of multiple music performance assessment conditions and suggested different ways for evaluating musical performance. McPherson and Thompson (1998) concluded that the purpose of the event strongly influenced the way a judge will listen to and evaluate a performance. Bergee and Platt (2003) found, "statistically significant differences in the main effects of time of day, type of event, and school size" (p. 342). McPherson and Thompson (1998) suggested that the size and acoustics of the performance space influenced music performance and assessment. These findings suggest that participants acknowledge these items' influence and consider them when making rating decisions.

#### **Factor Four—The Conductor's Contributions**

The fourth influence on adjudicators' rating of choral performance was the conductor's contributions. This factor accounted for 5.86% of the variance among the factors that influenced adjudicators' ratings. This finding, aspects of the conductor's contributions to the ensemble influencing adjudicators' perceptions and ratings of ensemble performance quality, is supported by prior research (Madsen, 2009; Morrison et al., 2009).

Items loaded, with correlation coefficients, on Factor Four (The Conductor's Contributions) were, in order: interpretation ( $r = .783$ ), musical style ( $r = .689$ ), and conducting gesture ( $r = .612$ ). Interpretation had a strong influence on Factor Four. Musical style and conducting gesture had a moderate influence on Factor Four.

When reviewing these items, the researcher deduced that they ultimately derive from the conductor. Although the first two items (Interpretation and Musical style) may be viewed as an ensemble characteristic, it is plausible that the conductor made musical and stylistic decisions unilaterally and coached the ensemble to perform what he or she wished. It is also plausible that the conductor and then ensemble collaborated during rehearsals to come to a unified interpretation, and agreed upon presentation, of the musical selection. In either scenario, the conductor influences the outcome substantively. The third item (Conducting gesture) is evidently a characteristic of the conductor. Traditionally, the conductor works with the ensemble to realize the conductor's vision of a musical performance. In doing so, the conductor sensitizes the ensemble to his or her conducting gestures. It is also likely that the conductor explores a repertoire of gestures with the ensemble to discover the most effective presentation that elicits the desired response. Ultimately, the conductor's aural image of the musical performance influences the musical decisions made by both the conductor and the ensemble. Factor Four is a result of the conductor's contributions.

In the present study, the conductor's contributions were found to influence adjudicators' rating decisions of choral ensemble performance. These findings are supported by prior research. Madsen (2009) concluded that different styles of conducting

behavior influenced audience perceptions of performance quality with poor conducting influencing ratings more than good conducting. Morrison et al. (2009) discovered that the more visually expressive the conductor appeared, the higher the ratings for the ensemble's performance. Results from the present study provide additional support that the conductor's physical gesture influences choral adjudicators' ratings of choral ensemble performance.

Choral adjudicators who participated in the present study recognized that the conductor plays a role in the presentation of a musical performance and that the conductor's contributions had an influence on adjudicators' ratings of performance. Participants in other studies mentioned that they only attended to audio information (sound) when assessing or evaluating a musical performance (Mitchell & MacDonald, 2016; Tsay, 2103), however; researchers determined that visual information complimented audio information and provided valuable insight regarding performer's intent and overall musical quality. The findings of the present study support other researchers' findings that the conductor has an influence on adjudicators' ratings (Madsen, 2009; Napoles, 2013).

Choral conductors can use the findings of the present study to inform their conducting behavior during adjudicated performance. When they are performing, their influence on the choir has an influence on adjudicators rating decisions of the overall performance. Professional music organization authorities can use these findings to support arguments regarding the use of blind adjudicated performance.

## **Research Question 2: Years of Adjudication Experience**

Research Question 2 was: “Do differences exist among adjudicators influences on rating decisions by years of adjudication experience?” A main effect was found for the factors—differences existed among the factors that influenced participants’ rating decisions of choral performance. There was neither a significant main effect for participants’ years of adjudication experience nor a significant interaction effect among the factors by participants’ years of adjudication experience. The null hypothesis, that no differences exist among adjudicators’ influences on rating decision by years of adjudication experience, failed to be rejected. Differences among factors of influence were not explained by participants’ years of adjudication experience.

The present study might be novel in that it focused on participants’ years of adjudication experience as ‘experience,’ specifically. The body of available research regarding music experience and its possible influence on judgments of music quality seemed to characterize experience as (a) formal music training and instruction and (b) experience teaching music. Although ‘adjudication experience’ was not the focus of her study, Mills (1991) argued that music educators’ years of adjudication experience provided enough training for end-of-year music performance assessment. Results from the present study suggest that years of adjudication experience might not have any influence on adjudicators’ rating decisions. These results are important because they might dispel assumptions regarding years of adjudication experience as a factor of influence on choral adjudicator rating decision.

### **Research Question 3: Academic Training**

Research question two was: “Do differences exist among adjudicators influences on rating decisions by academic training?” A main effect was found for the factors—differences existed among factors that influenced rating decisions of choral performance. There was neither a significant main effect for participants’ academic training nor a significant interaction effect among the factors by participants’ academic training. The null hypothesis, that no differences exist among adjudicators’ influences on rating decision by academic training, failed to be rejected. Differences among factors of influence were not explained by participants’ academic training.

The present study might be unique in that it focused on participants’ academic training as a possible influence on factors that influence choral performance ratings. For this study, academic training was defined as participants’ type of degree earned (degree focus) and participants were placed into three groups: (a) education, (b) performance, and (c) hybrid—education and performance. The body of available research regarding forms of training and their possible influence on judgments of music quality seemed to characterize training as (a) music experience, (b) music instruction, (c) teaching level, and (d) adjudicator training. The findings of the present study contribute to the available body of research to explain further possible areas of influence among music performance adjudicators.

Results from the present study suggest that academic training (degree focus) might not have an influence on adjudicators’ rating decisions of performance quality. These results are important because they may inform expectations regarding adjudicators’



ability to assess and evaluate choral music performance dependent on whether they hold degrees in music education, performance, or a blend of education and performance. These results might dispel possible assumptions and expectations regarding academic training as a factor of influence on choral adjudicators' rating decision.

### **Possible Methodological Limitations**

#### **Number of Participants**

The number of participants in the present study were limited by: (a) geographical boundaries (a delimitation imposed by the researcher), (b) state music organization representatives' willingness, ability, or availability to share contact information, (c) the accuracy of contact information, (d) effects of e-mail servers' security measures, and (e) participants' willingness and ability to volunteer.

#### **Available Data—Measure used to Collect Data**

The data used for this study was limited by the number of participants ( $N = 71$ ) who volunteered to take the CAPS. Data used for statistical analysis was limited due to the standard participant-to-variable ratio of three to one (3:1) (Asmus, 1989). Use of all 53 items would have required at least 159 participants (i.e., 53 items multiplied by 3). Due to the number of participants ( $N = 71$ ) used in the present study, a total of 23 items (i.e.,  $71/3$ ) with the highest commonality values were used as variables in the PCA. Ideally all 53 items from the CAPS would have been used as variables in the PCA, possibly resulting in more factors of influence on participants' rating decisions of choral performance. The use of more data might have produced different results from the two-way repeated measures ANOVA, possibly suggesting an interaction effect among factors

of influence with either participants' years of adjudication experience or academic training, or affirming the results of the present findings.

### **Lack of Prior Research on the Topic**

The scope of available published research on the topic of music performance assessment and evaluation is expansive; however, available literature investigating bias derived from participants' years of adjudication experience and academic training is limited. This limitation, of a limited number of studies, provides a less than ideal foundation of information upon which to draw. Nevertheless, this paucity of research revealed a gap within available literature and allowed for the present study to provide insight into possible bias among choral adjudicators stemming from (a) years of adjudication experience and (b) academic training.

### **Self-reported Data**

Data for the present study was collected via a five-point Likert-type scale in which participants' self-reported influences on decision-making processes. The prompt for each item on the scale was, "How strong of an influence is this item toward your rating decision when adjudicating a choral performance?" The possibility that a participant's responses might have been biased, in that they did not answer accurately (intentionally or unintentionally), must be considered. Bias responses might have been influenced by: (a) selective memory, (b) telescoping, (c) attribution, or (d) exaggeration. Selective memory is the act of remembering or not remembering an experience or event (e.g., a participant might recall a specific memory that could enhance, impair, or alter the content of the memory). Telescoping is the act of recalling an event that occurred at one

time as if it occurred at a different time (e.g., a participant recalls an event that influenced their behavior and perceives the event as occurring sooner or later than it actually did).

Attribution is the act of accrediting positive outcomes to one's own influence and negative outcomes to external forces (e.g., a participant responds to an item incorrectly when evaluating reasons for their own behaviors). Exaggeration is the act of embellishing an event or experience to make it seem more significant than it is in actuality (e.g., a participant increases or decreases the influence of an item beyond its actual influence on their behavior). Social desirability bias (SDB)—the participants' desire to edit responses in order to make him or herself look good—may also occur when using self-reported data (Gonyea, 2005). A combination of these bias conditions might have occurred, with or without participants' knowledge. Radocy's (1976) study on the effects of adjudicator bias determined that adjudicators were influenced by the opinions of those whom they viewed as authority figures. Nancarrow and Brace (2000) described participants who felt the need to preserve their self-esteem and present themselves favorably to the researcher. With Radocy's (1976), and Nancarrow and Brace's (2000) conclusions in mind, it is possible that the researcher influenced participants, or another figure perceived to be an authority in choral music, resulting in possible exaggerated responses or responses that the participants' thought would please the researcher.

Although the possibility exist that participants' self-reported responses might have been biased, the findings of the present study are valuable in steering research towards an understanding of influences on adjudicators' rating decisions and possible areas of bias that might interact with those influences.

Paulhus and Vazire (2007) wrote about the advantages of using self-reported data. They explained that self-reported data (a) provided rich information (e.g., no one has more intimate knowledge about the participants than the participants themselves), (b) motivated participants (e.g., in general, people are pleased to talk about themselves), (c) had a causal force (e.g., a measure that forces participants to reflect on self-perceptions of identity and influence on the world), and (d) was practical (e.g., an efficient and inexpensive way to collect data). Gonyea (2005) expressed that self-reported data is ‘attitudinal information’ and is “based on personal beliefs or perceptions” (p. 76). Researchers use of self-reported data with online surveys is pervasive in research practice (Tourangeau, 2004). Self-reported data collection methods are useful in that they enable participants to reflect on themselves and provide rich information regarding their beliefs and self-perceptions. Self-reported data is prevalent in research practice and is considered to be advantageous for constructing understanding about people (Paulhus & Vazire, 2007). Though ‘self-report’ as a method of data collection has its challenges, it also contributes positively to the knowledge base in music education and in other fields, and thus, it has a valid ‘home’ in social science research.

### **Possible Limitations of the Researcher**

#### **Access**

Data collection for the present study was effected by limited access to choral adjudicators’ contact information in some states within the Southeast United States. Choral activities chairpersons in eight of 11 states willingly provided contact lists of approved choral adjudicators, employed for NCMEA and ACDA sanctioned events. One

state was not permitted to share choral adjudicators' information due to its state music educator association's rules and regulations prohibiting the sharing of personal information. Two states did not respond to repeated requests for choral adjudicators' information.

### **Longitudinal Effects**

Data collection was limited due to time constraints set by the researcher. The CAPS was made available to participants for one month. It is possible that potential participants were unable to complete the survey during that time frame. It is possible that if the data collection instrument was available to potential participants for a longer period of time, more participants might have completed the scale.

### **Suggestions for Future Research**

The purpose of the present research study was to investigate influences on choral performance adjudicators' ratings decisions of choral performance quality and to determine if differences existed among participants by years of adjudication experience and academic training. The intent of the researcher was to focus on the adjudicator, the source of the rating during music performance adjudication. Rather than focusing on the performers, the performance environment, or the tools used to measure music performance, it was decided that the choral adjudicator needed to be the focus of investigation. McPherson and Thompson's (1998) Process Model of Music Performance Assessment provided a framework from which to focus areas of research. The fourth factor of the model, Evaluator Characteristics, seems to be a growing field in music research. More research is needed to further understand characteristics of the evaluator

and to investigate influences, and possible areas of bias, on adjudicators' rating decisions of choral performance quality. Researchers can use the present study as a foundation from which to move forward.

The present study provided specific information regarding (a) influences on adjudicators' ratings and (b) an explanation on possible differences based on years of adjudication experience and academic training. The current study suggests four factors of influence (i.e., the ensemble's performance, visual aspects, extra-musical aspects, and the conductor's contributions) on rating decisions. The present study suggests that years of adjudication experience and academic training may not contribute to the differences among adjudicators' ratings. Although these results were revealed within acceptable boundaries of quantitative statistical methods, these findings should be viewed with a degree of caution.

### **Replicating the Study**

Researchers should consider replicating the present study in other geographical areas to build a generous understanding of what influences adjudicators' rating decisions and investigate possible underpinnings of bias. More efforts might be made to acquire a more complete list of current choral performance adjudicators than was compiled in the present study. Researchers should work to secure accurate and reliable contact information for possible participants. Researchers might need to consider the effects of an email system's filtering programs for incoming emails. The present study was unable to reach some potential contacts due to email system filters that prevented possible spam emails from reaching the owner of the email address. Perhaps future researchers should

use traditional mail services in combination with electronic mail services when recruiting possible participants.

Future researchers should seek to increase the number of participants to gain a broader than current view of the influences on adjudicators' ratings and to determine if an interaction effect exists among those influences with participants' years of adjudication experience and academic training. Increasing the number of participants might aid in the study's ability to generalize results to the broader population. Increasing the number of participants might also allow for more items to be used in data collection, including items described in research literature regarding the influence of ethnicity, race, and gender (Brown & Novak, 2007; Cheek, 2007; Elliot, 1995/1996).

### **Perception Study**

Research using participants' self-reported data should be viewed with a degree of caution due to the possibility that participants' responses might not be precise and or fully accurate. To get a more complete and accurate understanding of the foundations of adjudicator's implicit bias and possible influences on rating decisions, future research is needed. Researchers might consider performing a perception study investigating implicit bias among choral performance adjudicators' influences on ratings. The current study provides a foundation for further inquiry into adjudicator bias.

### **Investigate Observations and Unexpected Findings from the Present Study**

Possible areas of study for future research might explore observations and findings that were unexpected in the current study. An investigation of differences among participants regarding factors of influence is suggested. The present study

revealed a significant main effect among the factors that influence adjudicators' rating decisions.

Participants' mean scores for Factor Three (extra-musical aspects) ranged from 'not strong' to 'somewhat strong' while participants' mean scores for Factor Four (the conductors' contribution) were 'strong.' It is interesting that Factor Three was ranked higher than Factor Four in the PCA while the mean scores reflect items with lower influence on participants' rating decisions. This unexpected observation, in the results of research questions one and two, may provide direction for future research.

Another unexpected discovery of the present study was the item 'ensemble's responsiveness to the conductor' and its influence on Factor Two (Visual Aspects). This item was not found in prior research and might suggest a new area for future research. Future researchers might explore this item in a variety of ways, most notably how this performance characteristic affects audience members' (and adjudicators') perceptions of music performance quality.

Future researchers might select to test the findings of the present PCA using a confirmatory factor analysis (CFA). Confirmatory factor analysis is used to test hypothesis and theories one expects to find regarding a set of variables (Vogt & Johnson, 2011). Due to the novelty of the researcher-generated CAPS measurement instrument, a CFA may be calculated to test its construct validity (Fabrigar & Wegener, 2012). Computing a CFA may provide further support for the factor structure revealed in the PCA (Besnoy, Dantzler, Besnoy, & Byrne, 2016).



An interesting observation of the present study was the order in which each item loaded on the factors in the PCA. Future researchers may consider calculating an item response theory (IRT) analysis to determine and explain possible relationships among variables (items) within each factor derived from the PCA of the present study.

According to Vogt and Johnson (2011), “The assumption is that each of the items is measuring some aspect of the same underlying (\*latent) ability, trait, or attitude.” (p. 189). Future researchers might find it useful to explore the relationships among items found within each factor in the present study.

### **Conclusion**

Even though related literature regarding music performance assessment was available prior to the present study, the present study was needed because of (a) expectations regarding music performance assessment, (b) growing concerns regarding bias that may influence adjudicators’ ratings of music performance, and (c) the scarcity of research investigating characteristics of the adjudicator, specifically. Accomplishing the purposes of the present research—to analyze influences on choral adjudicators’ rating decisions of choral performance and to determine if adjudicators’ years of adjudication experience and academic training (focus) had an effect on those influences—provides useful information for school music teachers, state-level music organization administrators, event coordinators, adjudicator training programs, as well as choral performance adjudicators themselves. It is the hope of the researcher that the findings from the present study, and the findings of any subsequent research related to the topic of

choral performance adjudication, are used to nurture and influence music teaching, learning, and performance.

## REFERENCES

- Alabama Vocal Association. (October, 2015). *Alabama Vocal Association Bylaws*.  
(Article IX, Section 9.1). Retrieved from [http://www.alavocal.org/public\\_pages/public\\_handbook\\_pages/Bylaws.pdf](http://www.alavocal.org/public_pages/public_handbook_pages/Bylaws.pdf)
- 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.
- Batey, A. L. (2002). Take the terror out of adjudication. *Teaching Music*, 10(3), 40–46.
- Batey, A. L. (2007). Preparing a high school choir for adjudication. *Choral Journal*, 48(6), 73–75.
- Besnoy, K. D., Dantzler, J., Besnoy, L. R., & Byrne, C. (2016). Using exploratory and confirmatory factor analysis to measure construct validity of the traits, aptitudes, and behaviors scale (TABS). *Journal for the Education of the Gifted*, 39(1), 3–22.
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51(2), 137–150. doi:10.2307/3345847
- Bergee, M. J. (2007). Performer, rater, occasion, and sequence as sources of variability in music performance assessment. *Journal of Research in Music Education*, 55(4), 344–358.
- Bergee, M. J., & McWhirter, J. L. (2005). Selected influences on solo and small-ensemble festival ratings: Replication and extension. *Journal of Research in Music Education*, 53(2), 177–190.

- 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*(4), 342–35.
- 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*(4), 358–374.
- Brown, A. S., & Novak, J. L. (2007). *Assessing the Intrinsic Impacts of a Live Performance*. San Francisco, CA: WolfBrown. Accessed online: <http://culturehive.co.uk/wp-content/uploads/2013/04/ImpactStudyFinalVersion.pdf>
- Burnsed, V., Hinkle, D., & King, S. (1985). Performance evaluation reliability at selected concert festivals. *Journal of Band Research, 22*(1), 22–29.
- Cassidy, J., & Sims, W. L. (1991). Effects of special education labels on peers' and adults' evaluations of a handicapped youth choir. *Journal of Research in Music Education, 39*(1), 23–34.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276.
- Child, D. (2006). *The essentials of factor analysis* (3rd ed.). New York, NY: Continuum International Publishing Group.
- Cheek, J. A. (2007). *The effect of race and racial perception on adjudicators' ratings of choral performances attributed to racially homogeneous and racially*

*heterogeneous groups* (Doctoral dissertation). Greensboro, NC: University of North Carolina at Greensboro.

Ciorba, C. R., & Smith, N. Y. (2009). Measurement of instrumental and vocal undergraduate performance juries using a multidimensional assessment rubric.

*Journal of Research in Music Education*, 57(1), 5–15.

Coleman, D. V. (1976). *Biographical, Personality, And Situational Determinants Of*

*Leisure Time Expenditure: With Specific Attention To Competitive Activities (athletics) And To More Cooperative Activities (music)* (Order No. 7614850).

Available from ProQuest Dissertations & Theses Global. (302789071).

Colver, M. C., & El-Alayli, A. (2016). Getting aesthetic chills from music: The

connection between openness to experience and frisson. *Psychology of Music*, 44(3), 413–427. Published online before print March 6, 2015.

doi:10.1177/0305735615572358

Cooksey, J. M. (1977). A facet-factorial approach to rating high school choral music

performance. *Journal of Research in Music Education*, 25(2), 100–114.

Corbin, L. A. (1995). Building a positive choral attitude. *Music Educators Journal*, 81(4),

24–49. doi:10.2307/3398832

Costa, P., & McCrae, R. (1992). *Revised NEO personality inventory (NEO-PI-R) and*

*NEO five-factor inventory (NEO-FFI) professional manual*. Odessa, FL:

Psychological Assessment Resources.

- Dalkey, N. C., & Helmer, O. (1963). An experimental application of the Delphi method to the user of experts. *Management Science*, *9*(3), 458–467. Retrieved from <http://www.jstor.org/stable/2627117>
- Daugherty, J. (1999). Spacing, formation, and choral sound: Preferences and perceptions of auditors and choristers. *Journal of Research in Music Education*, *47*(3), 224–238.
- DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11. Retrieved from <http://pareonline.net/getvn.asp?v=14&n=20>
- Duerksen, G. L. (1972). Some effects of expectation on evaluation of recorded musical performance. *Journal of Research in Music Education*, *20*(2), 268–272.
- Ekholm, E. (2000). The effect of singing mode and seating arrangement on choral blend and overall choral sound. *Journal of Research in Music Education*, *48*(2), 123–135.
- Elliot, C. (1995/1996). Race and gender as factors in judgments of musical performance. *Bulletin of the Council for Research in Music Education*, *Winter*(127), 50–56.
- Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. New York, NY: Oxford University Press.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *3*, 272–299.

- Florida School Music Association. (October, 2015). 2000-2001 State Music Performance Assessment Report. (Explanation of terms used in the state music report). Retrieved from <http://fsma.flmusiced.org/media/1128/mpa.pdf>
- Forbes, G. W. (1994). Evaluative music festivals and contests—are they fair? *Update: Applications of Research in Music Education*, 12(2), 16–20.  
doi: 10.1177/875512339401200203
- Ford, J. K. (1999). *The preference for strong or weak singer's formant resonance in choral tone quality* (Doctoral dissertation). Tallahassee, FL: The Florida State University.
- Fox, G. C. (1990). Making music festivals work. *Music Educators Journal*, 76(7), 59–64.
- Freer, P. (2011). The performance-pedagogy paradox in choral music teaching. *Music Faculty Publications*. Paper 38. [http://scholarworks.gsu.edu/music\\_facpub/38](http://scholarworks.gsu.edu/music_facpub/38)
- Geisler, C. D. (2014). *Evaluating music educator effectiveness through student growth: An investigation of the Ohio teacher evaluation system* (Doctoral dissertation). Minneapolis, MN: Capella University.
- Geringer, J. M., & Madsen, C. K. (1998). Musicians' ratings of good versus bad vocal and string performances. *Journal of Research in Music Education*, 46(4), 522–534.
- Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. *New Directions for Institutional Research*, 2005(127), 73–89.
- Gous-Kemp, C. S. (2014). The Delphi method for program enrichment: A variation on the theme. *Sosyal Bilimler Dergisi/Journal of Social Sciences*, 39(2), 235–242.

- Grant, J. W., & Norris, C. (1998). Choral music education: A survey of research 1982–1995. *Council for Research in Music Education, 135*, 21–59.
- Guegold, W. K. (1989). *An analysis of the adjudication results in the 1986-1988 Ohio Music Education Association state marching band finals with an emphasis on adjudicator consistency* (Order No. 9006128). Available from ProQuest Dissertations & Theses Global. (303788927).
- Hair, H. (1981). Verbal identification of music concepts. *Journal of Research in Music Education, 29*(1), 11–21.
- Haire, C. E. (2015). *Motivation for males to participate in private middle school and high school choruses* (Order No. 3708126). Available from Dissertations & Theses @ University of North Carolina at Greensboro. (1696782123).
- 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.  
doi:10.1177/0022429411434932
- Hewitt, M. P., & Smith, B. P. (2004). The influence of teaching-career level and primary performance instrument on the assessment of music performance. *Journal of Research in Music Education, 52*(4), 314–327.
- Howard, S. (2012). The effect of selected nonmusical factors on adjudicators' ratings of high school solo vocal performances. *Journal of Research in Music Education, 60*(2), 166–185.



- James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology, 69*, 85–98.
- Johnson, C. M. (1996). Musicians' and nonmusicians' assessment of perceived rubato in music performance. *Journal of Research in Music Education, 44*(1), 84–96.
- Jones, H. (1986). *An application of the facet-factorial approach to scale construction in the development of a rating scale for high school vocal solo performance (adjudication, evaluation, voice)*. Available from ProQuest Dissertations & Theses Global. (303512065).
- Juchniewicz, J. (2008). The influence of physical movement on the perception of musical performance. *Psychology of Music, 36*(4), 417–427.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychology Measurement, 20*, 141–151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 32–36.
- Kelly, G. (1955). *The Psychology of Personal Constructs*. New York, NY: Norton.
- Kinney, D. (2009). Internal consistency of performance evaluations as a function of music expertise and excerpt familiarity. *Journal of Research Music Education, 56*(4), 322–327.
- Krueger, H. (1966). A critical analysis of the Minnesota district and state band contest-festival system with implications for its improvement. *Journal of Band Research, 3*(1), 33–38.

- Laerd Statistics. (2015a). Principal components analysis (PCA) using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved from <https://statistics.laerd.com/>
- Laerd Statistics. (2015b). Two-way repeated measures ANOVA using SPSS Statistics. *Statistical tutorials and software guides*. Retrieved from <https://statistics.laerd.com/>
- Landy, F., & Farr, J. (1980). Performance rating. *Psychological Bulletin*, 87(1), 72–107.
- Larsen, R. J., & Diener, E. (1987). Affect intensity as an individual difference characteristic: A review. *Journal of Research in Personality*, 21, 1–39.
- Latimer, M. E. (2007). Adjudicator reliability: A comparison of the use of authentic state festival choir and global score audition forms. *Contributions to Music Education*, 34(1), 67–82.
- 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.  
doi:10.1177/0022429410369836
- Lehmann, A. C. (1997). Research note: Affective responses to everyday life events and music listening. *Psychology of Music*, 25(1), 84–90.
- Madsen, K. (2009). Effect of aural and visual presentation modes on Argentine and US musicians' evaluations of conducting and choral performance. *International Journal of Music Education*, 27(1), 48–58.

- Madsen, C. K., Geringer, J. M., & Wagner, M. J. (2007). Context specificity in music perception of musicians. *Psychology of Music, 35*(3), 441–451.
- McPherson, G. E. (1993). *Factors and abilities influencing the development of visual, aural and creative performance skills in music and their education implication (Doctoral Dissertation)*. University of Sydney, Australia.
- McPherson, G. E. (1995). The assessment of musical performance: Development and validation of five new measures. *Psychology of Music, 23*, 142–161.
- McPherson, G. E., & Thompson, W. F. (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education, 10*(1), 12–24.
- Merriam-Webster, Inc. (1983). *Webster's ninth new collegiate dictionary*. Springfield, MA: Merriam-Webster.
- Millard, E. M. (2014). *The role of ensemble competitions in choral music education (Doctoral dissertation)*. Tallahassee, FL: Florida State University. Available from ProQuest Dissertations & Theses Global. (1617975410).
- Mills, J. (1991). Assessing musical performance musically. *Educational Studies, 17*(2), 173–181.
- Mitchell, H. F., & MacDonald, R. A. R. (2016). What you see is what you hear: The importance of visual priming in music performer identification. *Psychology of Music*. Published online before print, February 10, 2016.  
doi:10.1177/0305735616628658
- Morgan, J., & Burrows, B. (1981). Sharpen your edge on choral competition. *Music Educators Journal, 67*(8), 44–47.

- 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*(1), 37–49.
- Nancarrow, C., & Brace, I. (2000). Saying the ‘right thing’: Coping with social desirability bias in marketing research. *Bristol Business School Teaching and Research Review, 3*, 1–11.
- Napoles, J. (2009). The effects of score use on musicians’ ratings of choral performance. *Journal of Research in Music Education, 57*(3), 267–279.
- Napoles, J. (2013). The influences of presentation modes and conducting gestures on the perceptions of expressive choral performance of high school musicians attending a summer choral camp. *International Journal of Music Education, 31*(3), 321–330.
- National Association for Music Education. (December, 2015). National Association for Music Education’s Structure [A description of the six geographical divisions of the professional organization]. Retrieved from <http://www.nafme.org/membership/educators/membership-tour/nafmes-structure/>
- Neil, R. J. (1945). *The development of the competition-festival in music education* (Order No. 0158251). Available from ProQuest Dissertations & Theses Global. (301826326).
- 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. doi:10.1177/002242940705500305

- North Carolina Music Educators Association. (2009). *NCMEA student event festivals and contests to change title* [Press release].
- North Carolina Music Educators Association. (March, 2015). *Policies and procedures*. (Section B.4.a). Retrieved from <http://www.ncmea.net/hs-choral-constitutionpolicies.html>
- Osborne, J. W., & Costello, A. B. (2009). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pan-Pacific Management Review, 12*(2), 131–146.
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, R. F. Krueger, R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York, NY: Guilford Press.
- Payne, B. (1997). A review of research on band competition. *Journal of Band Research, 33*(1), 1. Retrieved from <http://search.proquest.com/docview/1312111241?accountid=14604>
- Platz, F., & Kopiez, R. (2013). When the first impression counts: Music performers, audience and the evaluation of stage entrance behavior. *Musicae Scientiae, 17*(2), 167–197.
- Radocy, R. E. (1976). Effect of authority figure biases on changing judgments of musical events. *Journal of Research in Music Education, 24*, 119–128.
- Radocy, R. E. (1989). Evaluating student achievement. *Music Educators Journal, 76*(4), 30–33.

- Robinson, C. R. (1988). *Differentiated modes of choral performance evaluation using traditional procedures and a continuous response digital interface device* (Doctoral dissertation). Tallahassee, FL: Florida State University. Available from ProQuest Dissertations & Theses Global. (303554592).
- Robinson, C. R. (1990). Choral performance: Do you hear what I hear? *Music Educators Journal*, 77(4), 47–51. doi:10.2307/3397882
- Rohrer, T. (2002). The debate on competition in music in the twentieth century. *Update: Applications of Research in Music Education*, 21(1), 38–47.
- 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*. 42(2), 259–272.
- Selvey, J. D. (2014). *Visual and aural modes of perception in choral performance evaluations* (Doctoral dissertation). University of Washington: Seattle, WA. Available from ProQuest Dissertations & Theses Global. (1652865748).
- Silveira, J. M. (2014). The effect of body movement on listeners' perceptions of musicality in trombone quartet performance. *International Journal of Music Education*, 32(3), 311–323.
- Silvey, B., & Risher, R. (2015). Effects of conducting plane on band and choral musicians' perceptions of conductor and ensemble expressivity. *Journal of Research in Music Education*, 63(3), 369–383.

- Smith, B. P., & Barnes, G. V. (2007). Development and validation of an orchestra performance rating scale. *Journal of Research in Music Education, 55*(3), 268–280.
- South Carolina Music Educators Association. (October, 2015). *Choral division handbook*. (Solo and Ensemble Festival). Retrieved from <http://www.scmea.net/wp-content/uploads/2014/10/SCMEA-Choral-Division-Handbook-2015-2016.pdf>
- Springer, D. G., & Schlegel, A. L. (2016). Effects of applause magnitude and musical style on listeners' evaluations of wind band performance. *Psychology of Music, 44*(4), 742–756. doi:10.1177/0305735615591307
- Stanley, M., Brooker, R., & Gilbert, R. (2002). Examiner perceptions of using criteria in music performance assessment. *Research Studies in Music Education, 18*, 43–52.
- Stegman, S. F. (2009). Michigan state adjudicated choral festivals: Revising the adjudication process. *Music Educators Journal, 95*(4), 62–65.
- Stutheit, S. A. (1994). *Adjudicators', choral directors' and choral students' hierarchies of musical elements used in the preparation and evaluation of high school choral contest performance* (Doctoral dissertation). Kansas City, MO: University of Missouri. Available from ProQuest Dissertations & Theses Global. (304136400).
- Teachout, D. (1997). Preservice and experienced teachers' opinions of skills and behaviors important to successful music teaching. *Journal of Research in Music Education, 45*(1), 41–50.

- Teachout, D. J. (2004). Incentives and barriers for potential music teacher education doctoral students. *Journal of Research in Music Education, 52*, 234–247.  
doi:10.2307/3345857
- Teachout, D. J. (2008). Incentives and barriers for completing music teacher education doctoral programs. *Bulletin of the Council for Research in Music Education, 178*, 7–20.
- Thomas, K., Silvia, P., Nusbaum, E., Beaty, R., & Hodges, D. (2015). Openness to experience and auditory discrimination ability in music: An investment approach. *Psychology of Music*. Published online before print July 1, 2015.  
doi:10.1177/0305735615592013
- Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.
- Tourangeau, R. (2004). Survey research and societal change. *Annual Review of Psychology, 55*, 775–801.
- Tsay, C. (2013). Sight over sound in the judgment of music performance. *Proceedings of the National Academy of Sciences of the United States of America, 110*(36), 14580–14585.
- Vanweelden, K., & McGee, I. R. (2007). The influence of music style and conductor race on perceptions of choral ensemble and conductor performance. *International Journal of Music Education, 25*(1), 7–17.
- Vasconcellos, H. M. (2002). *The effect of choral performers' body movement on performance ratings assigned by high school choral students and college music*



- majors* (Doctoral dissertation). Kansas City, MO: University of Missouri.  
Available from ProQuest Dissertations & Theses Global. (305576607).
- Vogt, W. P., & Johnson, R. B. (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences*. Los Angeles, CA: Sage.
- Wagoner, C. L. (2011). *Defining and measuring music teacher identity: A study of self-efficacy and commitment among music teachers* (Order No. 3457596). Available from Dissertations & Theses @ University of North Carolina at Greensboro. (873786337).
- Ward, J. (2013). A preliminary study of the development of error detection and assessment skills in choral ensemble performances through adjudication training. In T. S. Brophy & A. Lehmann-Wermser (Eds.), *Music Assessment across cultures and continents: The culture of shared practice* (pp. 277–294). Chicago, IL: GIA Publications.
- Welch, G. F. (1994). The assessment of singing. *Psychology of Music*, 22(1), 3–19.
- Wrigley, W. J., & Emmerson, S. B. (2011). Ecological development and validation of a music performance rating scale for five instrument families. *Psychology of Music*, 41(1), 97–118.
- Zdzinski, S. F., & Barnes, G. V. (2002). Development and validation of a string performance rating scale. *Journal of Research in Music Education*, 50(3), 245–255.

**APPENDIX A****OPEN-ENDED QUESTIONNAIRE TO NORTH CAROLINA ADJUDICATORS**

What aspects do you consider when you adjudicate a choir performance at your state's music performance adjudication event? Please include as many aspects as you can think of in your response.

**APPENDIX B****STEP ONE:****LIST OF PERFORMANCE ASPECTS PROVIDED BY EIGHT RANDOMLY  
SELECTED NORTH CAROLINA ADJUDICATORS**

1. Use of the voice in a healthy manner
2. Use of the voice in a beautiful manner
3. Commitment to expressive qualities within the music
4. Commitment to a musical performance
5. Tone
6. Intonation
7. Diction
8. Choral technique
9. Vocal technique
10. Interpretation
11. Expression
12. Artistry
13. Aesthetics: Entrance
14. Aesthetics: Exit
15. Overall look
16. Demeanor
17. Focus of the group
18. Discipline of the group
19. Memorization
20. Blend
21. Use of dynamics
22. Use of implied dynamics
23. Expressiveness of singers
24. Adherence to standard performance practice of the musical selection
25. Voice placement
26. Forcing the voice
27. Pushing the voice
28. Good intonation
29. Vowel sounds
30. Clear consonants
31. Correct rhythms
32. Correct pitches
33. Use of correct etiquette
34. Look neat
35. Good first impression
36. Grade of music
37. Rehearsal time

38. General age of choir members
39. Department
40. Technical features in music
41. Use of phrasing (crescendos)
42. Proper pronunciation of words
43. Facial expression
44. Choral tone
45. Balance
46. Vowels
47. Stylistic performance
48. Suitability of music for the ensemble
49. Ensemble's conduct
50. Language
51. Ensemble issues
52. Is the ensemble together?
53. Is the phrasing together?
54. Accurate pitches
55. Accurate rhythms
56. Overall tone of ensemble
57. Breath support
58. Tone within sections
59. Tone across sections
60. Size of group
61. Sound like an ensemble
62. Adherence to musical instructions notated on music score
63. Musical performance
64. Tone quality
65. Intonation
66. Diction
67. Technical aspects of vocal production
68. Artistic style
69. Appearance
70. Dynamics (loud verses soft)
71. Expressive
72. Matching vowels
73. Consonants
74. Rehearsal time per week
75. Age level
76. Observe intentions of the composer
77. Music considerations

**APPENDIX C****STEP TWO:  
COMBINED LISTS OF PERFORMANCE ASPECTS**

- C.1 LIST OF PERFORMANCE ASPECTS FROM RELATED LITURATURE
- C.2 COMPLETE LIST OF PERFORMANCE ASPECTS FROM CHORAL  
ADJUDICTOR RESPONCES AND REALTED LITURATURE
- C.3 DISCRETE LIST OF PERFORMANCE ASPECTS WITH FREQUENCY  
COUN AND CATEGORY

## APPENDIX C.1

## LIST OF PERFORMANCE ASPECTS FROM RELATED LITURATURE

1. Accuracy of notes
2. Accuracy of rhythm
3. Appropriate tempo
4. Articulation
5. Assessment method
6. Balance
7. Blend
8. Body movement
9. Choral formation
10. Conducting
11. Conducting gesture
12. Conductor's race
13. Context
14. Control of medium
15. Degrees of movement
16. Diction
17. Dynamics
18. Effective dynamics
19. Ensemble
20. Ensemble expressivity
21. Expression
22. Expressiveness of the conductor
23. Expressivity
24. First impression
25. How the performer looks
26. Interpretation
27. Intonation
28. Location
29. Melodic accuracy
30. Movement
31. Musical effect
32. Musical style
33. Musicianship
34. Nature of the choir
35. Nature of the singer
36. Other performance factors
37. Overall performance
38. Performance attire
39. Phrasing

40. Precision
41. Prior knowledge regarding the performer
42. Rhythm
43. Rubato
44. Rhythmic accuracy
45. School size
46. School year/level
47. Sense of involvement in the music
48. Sense of performance
49. Singers' position (seating)
50. Skill
51. Soloistic singing
52. Spacing of singers
53. Stage deportment
54. Stage entrance
55. Style
56. Suitability
57. Suitable sense of style
58. Technique
59. Technique adequate to the piece
60. Tempo
61. Timbre
62. Time of day
63. Tone
64. Tone control
65. Tone quality
66. Type of event
67. Type of singing task
68. Use of musical score by adjudicator
69. Vibrato
70. Vocal production

## APPENDIX C.2

COMPLETE LIST OF PERFORMANCE ASPECTS FROM CHORAL ADJUDICATOR  
RESPONCES AND REALTED LITURATURE

1. Accuracy of notes
2. Accuracy of rhythm
3. Accurate pitches
4. Adherence to musical instructions notated on music score
5. Adherence to standard performance practice of the musical selection
6. Aesthetics: Entrance
7. Aesthetics: Exit
8. Age level
9. Appearance
10. Appropriate tempo
11. Articulation
12. Artistic style
13. Assessment method
14. Artistry
15. Balance
16. Blend
17. Body Movement
18. Breath support
19. Choral Formation
20. Choral Technique
21. Choral tone
22. Clear consonants
23. Commitment to a musical performance
24. Commitment to expressive qualities within the music
25. Conducting
26. Conducting gesture
27. Conductor's Race
28. Consonants
29. Context
30. Control of medium
31. Correct pitches
32. Correct Rhythms
33. Degrees of movement
34. Demeanor
35. Deportment
36. Diction
37. Discipline of the group
38. Dynamics



39. Effective dynamics
40. Ensemble
41. Ensemble Expressivity
42. Ensemble issues
43. Ensemble's conduct
44. Expression
45. Expressive
46. Expressiveness of singers
47. Expressiveness of the Conductor
48. Expressivity
49. Facial expression
50. First Impression
51. Focus of the group
52. Forcing the voice
53. General age of choir members
54. Good first impression
55. Good intonation
56. Grade of music
57. How the performer looks
58. Interpretation
59. Intonation
60. Is the ensemble together?
61. Is the phrasing together?
62. Language
63. Location
64. Look neat
65. Matching vowels
66. Memorization
67. Melodic accuracy
68. Movement
69. Musical considerations
70. Musical Effect
71. Musical Performance
72. Musical Style
73. Musicianship
74. Nature of the choir
75. Nature of the singer
76. Observe intentions of the composer
77. Other performance factors
78. Overall performance
79. Overall look
80. Overall tone of ensemble
81. Performance attire
82. Phrasing

83. Precision
84. Prior Knowledge regarding the performer
85. Proper pronunciation of words
86. Pushing the voice
87. Rehearsal time
88. Rehearsal time per week
89. Rhythm
90. Rubato
91. Rhythmic accuracy
92. School Size
93. School Year/Level
94. Sense of involvement in the music
95. Sense of performance
96. Singers' position (Seating)
97. Size of group
98. Skill
99. Soloistic Singing
100. Sound like an ensemble
101. Spacing of singers
102. Stage deportment
103. Stage Entrance
104. Style
105. Stylistic performance
106. Suitability
107. Suitability of music for the ensemble
108. Suitable sense of style
109. Technical aspects of vocal production
110. Technical features in music
111. Technique
112. Technique adequate to the piece
113. Tempo
114. Timbre (Tone color)
115. Time of day
116. Tone
117. Tone across sections
118. Tone Control
119. Tone Quality
120. Tone within sections
121. Type of Event
122. Type of singing task
123. Use of correct etiquette
124. Use of dynamics
125. Use of implied dynamics
126. Use of musical score by adjudicator

127. Use of phrasing (crescendos)
128. Use of voice in a beautiful manner
129. Use of voice in a healthy manner
130. Vibrato
131. Vocal Production
132. Vocal Technique
133. Voice placement
134. Vowel sounds
135. Vowels

## APPENDIX C.3

DISCRETE LIST OF PERFORMANCE ASPECTS WITH FREQUENCY COUNT  
AND CATEGORY

<b>Item Description</b>	<b>Freq.</b>	<b>Category</b>
Tone quality	10	Tone/vocal production
Vocal technique	10	Vocal technique
Dynamics	9	Technical accuracy
Musical expression	9	Interpretation
Blend	8	Tone/vocal production
Diction	7	Vocal technique
Harmonic intonation	7	Intonation (technical accuracy)
Interpretation	7	Interpretation
Adherence to musical instructions notated on music score	6	Technical accuracy
Musical style	6	Interpretation
Accuracy of rhythm	5	Technical accuracy
Adherence to rules of conduct and behavior for the event.	5	Extra-musical: Stage presence
Appearance	5	Extra-musical: Stage presence
Accuracy of pitches	4	Technical accuracy
Appropriateness of consonants	4	Vocal technique
Appropriateness of tempo	4	Interpretation
Appropriateness of vowels	4	Vocal technique
Balance	4	Tone/vocal production
Ensemble movement	4	Extra-musical: Stage presence
Age level	3	Extra-musical: Demographics
Conducting gesture	3	Conductor
Context of the performance event (purpose and location)	3	Extra-musical: Performance details
Method used to evaluate musical performance (i.e., multidimensional rubric, holistic, etc.)	3	Scoring
Precision	3	Technical accuracy
Standing formation of the group	3	Extra-musical: Stage presence
Appropriateness of music for the ensemble	2	Extra-musical: Performance details
Ensemble Size	2	Extra-musical: Demographics

<b>Item Description</b>	<b>Freq.</b>	<b>Category</b>
First Impression	2	Extra-musical: Stage presence
How the ensemble walks onto the stage	2	Extra-musical: Stage presence
Knowledge regarding the amount of time the choral ensembles rehearsed prior to the performance.	2	Extra-musical: Demographics
Phrasing	2	Interpretation
Tone control	2	Tone/vocal production
Adherence to standard performance practice of the musical selection	1	Interpretation
Appropriateness of vibrato	1	Interpretation
Attentiveness to the conductor	1	Ensembleship (ability to function as a group, musically)
Audience response	1	Extra-musical: Performance details
Body alignment	1	Vocal technique
Breath support	1	Vocal technique
Difficulty level of music	1	Extra-musical: Performance details
Ensemble's commitment to a musical performance	1	Extra-musical: Stage presence
Ethnicity	1	Extra-musical: Demographics
Expressiveness of the conductor	1	Conductor
Facial Expression	1	Extra-musical: Stage presence
How the ensemble walks off the stage	1	Extra-musical: Stage presence
Melodic Intonation	1	Intonation (technical accuracy)
Memorization	1	Performance expectation
Musicianship of the ensemble	1	Ensembleship (ability to function as a group, musically)
Performance space	1	Extra-musical: Performance details
Prior Knowledge regarding the performers	1	Extra-musical: demographics
Rubato	1	Interpretation
School Size	1	Extra-musical: demographics
Sense of Ensemble	1	Ensembleship (ability to function as a group, musically)
Time of day	1	Extra-musical: Performance details

## APPENDIX D

### PILOT SURVEY SENT TO PANEL OF EXPERTS

#### Choral Adjudicators' Preference Scale

The following items are designed to gather information about aspects that would influence (or have influenced) your rating decision when adjudicating a choral performance. For each item, please indicate its strength as an influence by circling one number on the five-point scale:

(1 = not strong, 2 = somewhat strong, 3 = strong, 4 = very strong, 5 = extremely strong)

ITEMS	CATEGORY	How strong of <b>an influence</b> is this item toward your rating decision when adjudicating a choral performance?				
		Not Strong	Somewhat Strong	Strong	Very Strong	Extremely Strong
1) Conducting gesture	Conductor	1	2	3	4	5
2) Expressiveness of the Conductor	Conductor	1	2	3	4	5
3) Attentiveness to the conductor	Ensembleship (ability to function as a group, musically)	1	2	3	4	5
4) Musicianship of the Ensemble	Ensembleship (ability to function as a group, musically)	1	2	3	4	5
5) Sense of Ensemble	Ensembleship (ability to function as a group, musically)	1	2	3	4	5
6) Age level	Extramusical: Demographics	1	2	3	4	5
7) Ensemble Size	Extramusical: Demographics	1	2	3	4	5
8) Ethnicity	Extramusical: Demographics	1	2	3	4	5
9) Knowledge regarding the amount of time the choral ensembles rehearsed prior to the performance.	Extramusical: Demographics	1	2	3	4	5
10) Prior Knowledge regarding the performers	Extramusical: Demographics	1	2	3	4	5
11) School Size	Extramusical: Demographics	1	2	3	4	5
12) Appropriateness of music for the ensemble	Extramusical: Performance Details	1	2	3	4	5
13) Audience response	Extramusical: Performance Details	1	2	3	4	5

14) Context of the performance event (purpose and location)	Extramusical: Performance Details	1	2	3	4	5
15) Difficulty level of music	Extramusical: Performance Details	1	2	3	4	5
16) Performance Space	Extramusical: Performance Details	1	2	3	4	5
17) Time of day	Extramusical: Performance Details	1	2	3	4	5
18) Adherence to rules of conduct and behavior for the event.	Extramusical: Stage Presence	1	2	3	4	5
19) Appearance	Extramusical: Stage Presence	1	2	3	4	5
20) Ensemble's commitment to a musical performance	Extramusical: Stage Presence	1	2	3	4	5
21) Ensemble movement	Extramusical: Stage Presence	1	2	3	4	5
22) Facial Expression	Extramusical: Stage Presence	1	2	3	4	5
23) First Impression	Extramusical: Stage Presence	1	2	3	4	5
24) How the ensemble walks onto the stage	Extramusical: Stage Presence	1	2	3	4	5
25) How the ensemble walks off the stage	Extramusical: Stage Presence	1	2	3	4	5
26) Standing formation of the group	Extramusical: Stage Presence	1	2	3	4	5
27) Adherence to standard performance practice of the musical selection	Interpretation	1	2	3	4	5
28) Appropriateness of tempo	Interpretation	1	2	3	4	5
29) Appropriateness of vibrato	Interpretation	1	2	3	4	5
30) Interpretation	Interpretation	1	2	3	4	5
31) Musical Expression	Interpretation	1	2	3	4	5
32) Musical Style	Interpretation	1	2	3	4	5
33) Phrasing	Interpretation	1	2	3	4	5
34) Rubato	Interpretation	1	2	3	4	5
35) Harmonic Intonation	Intonation (Technical Accuracy)	1	2	3	4	5

36) Melodic Intonation	Intonation (Technical Accuracy)	1	2	3	4	5
37) Memorization	Performance Expectation	1	2	3	4	5
38) Method used to evaluate musical performance (i.e. multidimensional rubric, holistic, ect.)	Scoring	1	2	3	4	5
39) Accuracy of pitches	Technical Accuracy	1	2	3	4	5
40) Accuracy of rhythm	Technical Accuracy	1	2	3	4	5
41) Adherence to musical instructions notated on music score	Technical Accuracy	1	2	3	4	5
42) Dynamics	Technical Accuracy	1	2	3	4	5
43) Precision	Technical Accuracy	1	2	3	4	5
44) Balance	Tone/Vocal Production	1	2	3	4	5
45) Blend	Tone/Vocal Production	1	2	3	4	5
46) Tone Control	Tone/Vocal Production	1	2	3	4	5
47) Tone Quality	Tone/Vocal Production	1	2	3	4	5
48) Appropriateness of Consonants	Vocal Technique	1	2	3	4	5
49) Appropriateness of Vowels	Vocal Technique	1	2	3	4	5
50) Body Alignment	Vocal Technique	1	2	3	4	5
51) Breath Support	Vocal Technique	1	2	3	4	5
52) Diction	Vocal Technique	1	2	3	4	5
53) Vocal Technique	Vocal Technique	1	2	3	4	5



**APPENDIX E****CHORAL ADJUDICATOR PREFERENCE SCALE (CAPS)**

<p>Choral Adjudicator Preference Scale</p> <p>Demographic Section</p> <p>How many years have you been teaching, including the current school year?</p> <ul style="list-style-type: none"><li><input type="radio"/> Slider Bar (1-50)</li></ul> <p>What best reflects your current primary school location?</p> <ul style="list-style-type: none"><li><input type="radio"/> Urban</li><li><input type="radio"/> Suburban</li><li><input type="radio"/> Rural</li></ul> <p>What best reflects the level of your current priority school assignment (or previous school if retired)?</p> <ul style="list-style-type: none"><li><input type="radio"/> K-5</li><li><input type="radio"/> 6-8</li><li><input type="radio"/> 9-12</li><li><input type="radio"/> Higher Education (University and College Level)</li></ul> <p>How many years have you been adjudicating choral performance, including the current school year?</p> <ul style="list-style-type: none"><li><input type="radio"/> Slider Bar (1-50).</li></ul> <p>Please list academic degrees earned as well as their concentration (e.g. Bachelor Degree, Music Education; Master Degree, Vocal Performance; Doctorate in the Musical Arts, Conducting Performance)</p> <ul style="list-style-type: none"><li><input type="radio"/> TEXT BOX</li></ul> <p>What is your age?</p> <ul style="list-style-type: none"><li><input type="radio"/> Slider Bar (20-80)</li></ul> <p>What is your gender?</p> <ul style="list-style-type: none"><li><input type="radio"/> Male</li><li><input type="radio"/> Female</li><li><input type="radio"/> Other</li></ul>
---

## Survey Section

The following items are designed to gather information about aspects that would influence (or have influenced) your rating decision when adjudicating a choral performance. For each item, please indicate its strength as an influence by selecting one number on the five-point scale:

(1 = not strong, 2 = somewhat strong, 3 = strong, 4 = very strong, 5 = extremely strong)

ITEMS	How strong of <b>an influence</b> is this item toward your rating decision when adjudicating a choral performance?				
	Not Strong	Somewhat Strong	Strong	Very Strong	Extremely Strong
Facial expression of the ensemble members	1	2	3	4	5
Knowledge regarding the amount of time the choral ensemble rehearsed prior to the performance	1	2	3	4	5
Interpretation	1	2	3	4	5
Body alignment	1	2	3	4	5
Appropriateness of tempo	1	2	3	4	5
Tone quality	1	2	3	4	5
Time of day	1	2	3	4	5
Method used to evaluate musical performance (e.g. multidimensional rubric, holistic, etc.)	1	2	3	4	5
Dynamics	1	2	3	4	5
Appropriateness of vibrato	1	2	3	4	5
Accuracy of pitches	1	2	3	4	5
Sense of ensemble	1	2	3	4	5
Expressiveness of the conductor	1	2	3	4	5
Diction	1	2	3	4	5
Audience response	1	2	3	4	5
Appropriateness of consonants	1	2	3	4	5
Balance	1	2	3	4	5
Memorization	1	2	3	4	5
Rubato	1	2	3	4	5
Precision	1	2	3	4	5
Appearance	1	2	3	4	5
Musical expression	1	2	3	4	5
Context of the performance event (purpose and location)	1	2	3	4	5
Standing formation of the group	1	2	3	4	5
School size	1	2	3	4	5
Tone control	1	2	3	4	5
How the ensemble walks off the stage	1	2	3	4	5
Harmonic intonation	1	2	3	4	5
Accuracy of rhythm	1	2	3	4	5
The degree to which the accompanist and the ensemble work together to enhance the musical performance	1	2	3	4	5
Appropriateness of music for the ensemble	1	2	3	4	5

Age level	1	2	3	4	5
Difficulty level of music	1	2	3	4	5
Phrasing	1	2	3	4	5
Adherence to musical instructions notated on music score	1	2	3	4	5
Conducting gesture	1	2	3	4	5
Ensemble's responsiveness to the conductor	1	2	3	4	5
Vocal technique	1	2	3	4	5
How the ensemble walks onto the stage	1	2	3	4	5
Appropriateness of vowels	1	2	3	4	5
Performance space	1	2	3	4	5
Musical style	1	2	3	4	5
Ensemble size	1	2	3	4	5
Melodic intonation	1	2	3	4	5
First impression	1	2	3	4	5
Musicianship of the ensemble	1	2	3	4	5
Prior knowledge regarding the performers	1	2	3	4	5
Ensemble's commitment to a musical performance	1	2	3	4	5
Breath support	1	2	3	4	5
Blend	1	2	3	4	5
Adherence to rules of conduct and behavior for the event	1	2	3	4	5
Adherence to standard performance practice of the musical selection	1	2	3	4	5
Ensemble movement	1	2	3	4	5

**APPENDIX F**

**INSTITUTIONAL REVIEW BOARD NOTIFICATION**

F.1 INSTITUTIONAL REVIEW BOARD NOTICE

F.2 INSTITUTIONAL REVIEW BOARD MODIFICATION NOTICE

## APPENDIX F.1

## INSTITUTIONAL REVIEW BOARD NOTICE

IRB <ori@uncg.edu>

to me, cifarrio, irbcorre, djteacho

2/4/16



To: Christopher Hansen  
Music Education

From: UNCG IRB

Date: 2/04/2016

RE: Determination that Research or Research-Like Activity does not require IRB Approval  
Study #: 16-0017

**Study Title:** An Analysis of Choral Performance Adjudicators' Values Related to Choral Performance

This submission was reviewed by the above-referenced IRB. The IRB has determined that this submission does not constitute human subjects research as defined under federal regulations [45 CFR 46.102 (d or f)] and does not require IRB approval.

**Study Description:**

The purpose of the study is to determine what choral music adjudicators' value about the quality of a choral ensemble's performance. Music performance adjudication is a common experience for choral teachers and students, and music educators have expressed concerns related to adjudicator bias . By examining music performance adjudicators' expectations of choral performance, the researcher hopes to learn how adjudicators arrive at their judgments. A modified Delphi method will be used.

- **When the survey is developed, please submit the survey questions as a modification to ensure the study continues to meet this determination.**

If your study protocol changes in such a way that this determination will no longer apply, you should contact the above IRB before making the changes.

CC:

David Teachout, Univ Tchg and Learning Commons

## APPENDIX F.2

## INSTITUTIONAL REVIEW BOARD MODIFICATION NOTICE

---

IRB <ori@uncg.edu>

9/29/16



🔒 to me, cifarrio, irbcorre, djteacho ▾

To: Christopher Hansen  
Music Education

From: UNCG IRB

Date: 9/29/2016

RE: Determination that Research or Research-Like Activity does not require IRB Approval  
Study #: 16-0017 (modification)

**Study Title:** An Analysis of Choral Performance Adjudicators' Values Related to Choral Performance

This submission was reviewed by the above-referenced IRB. The IRB has determined that this study continues to not constitute human subjects research as defined under federal regulations [45 CFR 46.102 (d or f)] and does not require IRB approval.

**Study Description:**

The purpose of the study is to determine what choral music adjudicators' value about the quality of a choral ensemble's performance. Music performance adjudication is a common experience for choral teachers and students, and music educators have expressed concerns related to adjudicator bias. By examining music performance adjudicators' expectations of choral performance, the researcher hopes to learn how adjudicators arrive at their judgments. A modified Delphi method will be used.

If your study protocol changes in such a way that this determination will no longer apply, you should contact the above IRB before making the changes.

CC:

David Teachout, Univ Tchg and Learning Commons

## APPENDIX G

### RECRUITMENT LETTER TO RANDOMLY SELECTED NORTH CAROLINA CHORAL PERFORMANMCE ADJUDICATORS

#### RECRUITMENT LETTER TO NC ADJUDICATORS

Dear North Carolina Choral Performance Adjudicator,

I am Christopher Hansen, a PhD student at the University of North Carolina at Greensboro, and the former Choral Activities Coordinator for the North Carolina Music Educators Association. I am contacting you to ask if you would be willing to participate in my dissertation research project. In my dissertation research project, I will attempt to describe what choral adjudicators listen for and value in a choral performance.

Prior researchers focused on the adjudication process, the adjudication environment, or adjudication instrument—the ratings rubric or ratings form. These researchers worked to increase the reliability of the adjudicators' ratings through controlling the performance setting or improving the ratings rubric, however; statistical analysis continues to show us that the reliability, or agreement among adjudicators, regardless of efforts for improvement, continues to be far less than ideal. My research will not focus on the adjudication environment or the adjudication instrument but rather my research will focus on the adjudicator to learn what he or she values most in a choral performance.

If you choose to participate I will send you a participant consent form to complete. Once the consent form is completed, you will be directed to a short online survey. The online survey will be in the form of an open-ended question regarding the aspects you consider when you adjudicate choral performances at your state's music performance adjudication events. The survey will require you to reflect on your experiences as an adjudicator at choral performance events. The duration of your participation in this project will be limited to the completion of the survey.

If you have any questions concerning this research study, please contact me at 910-280-4865, or [choirnrd@gmail.com](mailto:choirnrd@gmail.com). Thank you for considering participating in this important music education research.

Sincerely,

Chris Hansen

## APPENDIX H

### CAPS PARTICIPANT RECRUITMENT LETTER

#### RECRUITMENT LETTER TO CHORAL ADJUDICATORS

Dear Choral Adjudicator,

I am Christopher Hansen, a PhD student at the University of North Carolina at Greensboro, and the former Choral Activities Coordinator for the North Carolina Music Educators Association. I am contacting you to ask if you would be willing to participate in my dissertation research project. In my dissertation research project, I will attempt to describe what choral adjudicators listen for and value in a choral performance.

Prior researchers focused on the adjudication process, the adjudication environment, or adjudication instrument—the ratings rubric or ratings form. These researchers worked to increase the reliability of the adjudicators' ratings through controlling the performance setting or improving the ratings rubric, however; statistical analysis continues to show us that the reliability, or agreement among adjudicators, regardless of efforts for improvement, continues to be far less than ideal. My research will not focus on the adjudication environment or the adjudication instrument but rather my research will focus on the adjudicator to learn what he or she values most in a choral performance.

If you choose to participate I will send you a participant consent form to complete. Once the consent form is completed, you will be directed to a short online survey. You will be guided through the process electronically. The survey will require you to reflect on your experiences as an adjudicator at choral performance events. The duration of your participation in this project will be limited to the completion of the survey.

If you have any questions concerning this research study, please contact me at 910-280-4865, or [choirnrd@gmail.com](mailto:choirnrd@gmail.com). Thank you for considering participating in this important music education research.

Sincerely,

Chris Hansen



## APPENDIX I

### CONSENT FORM

#### UNIVERSITY OF NORTH CAROLINA AT GREENSBORO CONSENT TO ACT AS A HUMAN PARTICIPANT

Project Title: An Analysis of Choral Performance Adjudicator's Values Related to Choral Performance

Principal Investigator and Faculty Advisor: Christopher Hansen/Dr. David Teachout.

Participant's Name: \_\_\_\_\_

**What are some general things you should know about research studies?**

You are being asked to take part in a research study. Your participation in the study is voluntary. You may choose not to join, or you may withdraw your consent to be in the study, for any reason, without penalty.

Research studies are designed to obtain new knowledge. This new information may help people in the future. There may not be any direct benefit to you for being in the research study. There also may be risks to being in research studies. If you choose not to be in the study or leave the study before it is done, it will not affect your relationship with the researcher or the University of North Carolina at Greensboro. Details about this study are discussed in this consent form. It is important that you understand this information so that you can make an informed choice about being in this research study.

You will be given a copy of this consent form. If you have any questions about this study at any time, you should ask the researchers named in this consent form. Their contact information is below.

**What is the study about?**

This is a research project. Your participation is voluntary. The purpose of this research project is to determine what choral music adjudicators' value about the quality of a choral ensemble's performance. Music performance adjudication is a common experience for choral teachers and students. By examining music performance adjudicators' expectations of choral performance, the researcher hopes to learn how adjudicators arrive at their judgments.

**Why are you asking me?**

You are invited to participate in this research project because you are a choral performance adjudicator within the Southern Division of the National Association for Music Education (NAfME). Only choral performance adjudicators within the Southern Division of NAfME are eligible to participate.

**What will you ask me to do if I agree to be in the study?**

You will be given a short survey to complete. The survey will address what aspects you consider when you adjudicate a choir performance at your state's music performance adjudication event. The survey will be provided through an online survey program—Qualtrics. You will be guided through the process electronically. The survey will require you to reflect on your experiences as an adjudicator at choral performance events. The duration of your participation is limited to the completion of the survey. If you have any questions related to the survey, or to aid your full understanding of what you are consenting to, please contact Mr. Christopher Hansen, primary investigator, at 910-280-4865.

**What are the risks to me?**

The Institutional Review Board at the University of North Carolina at Greensboro has determined that participation in this study poses minimal risk to participants.

If you have questions, want more information or have suggestions, please contact Christopher Hansen at 910-280-4865 or [choimrd@gmail.com](mailto:choimrd@gmail.com) or contact Dr. David Teachout at 336-334-4759 or [djteacho@uncg.edu](mailto:djteacho@uncg.edu). If

you have any concerns about your rights, how you are being treated, concerns or complaints about this project or benefits or risks associated with being in this study please contact the Office of Research Integrity at UNCG toll-free at (855)-251-2351.

**Are there any benefits to society as a result of me taking part in this research?**

This research seeks to identify what choral performance adjudicators' value regarding the quality of choral ensembles' performance. A general understanding among high school chorus teachers and program stakeholders is that music performance evaluations are intended to be positive educational experiences for both students and teachers. For many choral music programs the reality of a festival or competition is perceived to be the ultimate measure of an ensembles performance quality. Administrators, program stakeholders, and teachers see state music performance adjudication as synonymous to a state standardized test, using the rating as a means to represent the quality of teaching and learning by comparing the rating to the approved standard. Music education advocates view these musical performance evaluations as a means to justify the importance of music education in schools.

Researchers explored the topic of music performance adjudication and adjudicator training. They found evidence for concern regarding reliability among adjudicator's evaluations of choral performance. They suggested that adjudicator training might improve adjudicator reliability. For decades, music education researchers sought to improve music performance evaluation practices by analyzing the measurement instrument on evaluation rubrics and ratings forms. To better understand why adjudicators make the decisions they make, that which they value in a music performance should be explored. Identifying and acknowledging adjudicators' values during their perceptions of a musical performance may have positive implications for future adjudication training programs, for state-mandated music performance evaluation processes in general, and choral singers and teachers performance experiences.

**Are there any benefits to *me* for taking part in this research study?**

There are no direct benefits to participants in this study.

**Will I get paid for being in the study? Will it cost me anything?**

There are no costs to you or payments made for participating in this study.

**How will you keep my information confidential?**

The data collection procedure will be confidential. Study participants will not put their names on their electronic surveys, and their names will therefore not be associated with their responses. The completed surveys will be stored electronically and will be password protected. All information obtained in this study is strictly confidential unless disclosure is required by law. Absolute confidentiality of data provided through the Internet cannot be guaranteed due to the limited protections of Internet access. Please be sure to close your browser when finished so no one will be able to see what you have been doing.

**What if I want to leave the study?**

You have the right to refuse to participate or to withdraw at any time, without penalty. If you do withdraw, it will not affect you in any way. If you choose to withdraw, you may request that any of your data which has been collected be destroyed unless it is in a de-identifiable state. The investigators also have the right to stop your participation at any time. This could be because you have had an unexpected reaction, or have failed to follow instructions, or because the entire study has been stopped.

**What about new information/changes in the study?**

If significant new information relating to the study becomes available which may relate to your

willingness to continue to participate, this information will be provided to you.

**Voluntary Consent by Participant:**

By signing this consent form/completing this survey/activity (used for an IRB-approved waiver of signature) you are agreeing that you read, or it has been read to you, and you fully understand the contents of this document and are openly willing consent to take part in this study. All of your questions concerning this study have been answered. By signing this form, you are agreeing that you are 18 years of age or older and are agreeing to participate, or have the individual specified above as a participant participate, in this study described to you by Christopher Hansen.

Signature: \_\_\_\_\_ Date: \_\_\_\_\_