

Spring 2020

## **Analysis of a Research-Based Show Choir Competition Adjudication Rubric: Reliability and User Perceptions**

LaDona Tyson

Follow this and additional works at: <https://aquila.usm.edu/dissertations>



Part of the [Music Education Commons](#)

---

### **Recommended Citation**

Tyson, LaDona, "Analysis of a Research-Based Show Choir Competition Adjudication Rubric: Reliability and User Perceptions" (2020). *Dissertations*. 1784.  
<https://aquila.usm.edu/dissertations/1784>

This Dissertation is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Dissertations by an authorized administrator of The Aquila Digital Community. For more information, please contact [Joshua.Cromwell@usm.edu](mailto:Joshua.Cromwell@usm.edu).

ANALYSIS OF A RESEARCH-BASED SHOW CHOIR COMPETITION  
ADJUDICATION RUBRIC:  
RELIABILITY AND USER PERCEPTIONS

by

LaDona K. Tyson

A Dissertation  
Submitted to the Graduate School,  
the College of Arts and Sciences  
and the School of Music  
at The University of Southern Mississippi  
in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy

Approved by:

Dr. Gregory Fuller, Committee Chair  
Dr. Mark Malone  
Dr. Holly Foster  
Dr. Edward Hafer  
Dr. Webb Parker

---

Dr. Gregory Fuller  
Committee Chair

---

Dr. Jay Dean  
Director of School

---

Dr. Karen S. Coats  
Dean of the Graduate School

May 2020

COPYRIGHT BY

LaDona K. Tyson

2020

*Published by the Graduate School*



## ABSTRACT

A weighted rubric, created in a prior pilot study, used for adjudicating nine show choir competitions was statistically analyzed to determine reliability. A survey and ten interviews with adjudicators were examined to ascertain user perceptions. Results from an intraclass correlation coefficient indicate that the Tyson Rubric has excellent inter-rater reliability. Survey respondents indicated that the Tyson Rubric provides a step towards standardization in the genre of show choir, though improvements were suggested. Interviews were transcribed and coded, revealing themes of bias in show choir adjudication, the need for standardization in the genre, and the need for improved adjudication training. The study suggests that further research is needed to determine better practices in show choir adjudication due to a lack of empirical research in the genre.

## ACKNOWLEDGMENTS

I would first like to thank the instructors and mentors who guided this process with patience, support, and encouragement. My committee chair, Dr. Gregory Fuller, has been instrumental to this process, pushing me to put words on paper, and always taking time to mentor and advise, despite his busy schedule. Thank you to Dr. Holly Foster for a detailed eye and quick editing feedback, making the process easier and less intimidating. I would like to thank Dr. Ed Hafer for graciously taking time to answer questions and provide support, along with stellar travel advice. Great credit goes to Dr. Mark Malone for the endless hours spent reading, editing, and discussing every word of the document, and for his encouragement and positivity along the way. Finally, thank you to Dr. Webb Parker for believing in me enough to recruit me to the program, and for sticking with me during the most difficult times (#RIPReimer).

This project would not have been possible without assistance from the show choir directors in Mississippi: Shanna Luckett, Debbie Wolfe, Amy Whittenton, Shane Cockrell, Chris Word, Ashley Bolling, Mitchell McGinnis, Toney Keeler, and Mady McNemar. Also, thanks to Sherman Charles and Carmen Scoring Systems.

Thank you to a host of friends and family who encouraged me throughout the journey, but specifically to Jonathan Kilgore, who listened patiently to every complaint and equally celebrated every victory.

## DEDICATION

I would like to dedicate this dissertation to my family. I am a first-generation college student, with parents who worked tirelessly to give me opportunities they did not have. This dissertation is dedicated to Dexter and Joyce Ferguson, who supported every step in my journey and gave me the opportunity to dream big and achieve. I owe all of my success to the foundation you gave me. While I'm sad that you are not here with me during this time, Dad, I know that you are watching and are proud. Mom – nothing I do is without your love and support. I got my strength from you.

Also, I would like to dedicate this work to my husband and daughters. Thank you to Scott Tyson for never letting me quit and for his quiet, yet strong, voice of encouragement along the way. Spending life with you is my greatest joy. This degree was a team effort and I could never have finished without you. Finally, to Claire and Norah - though I was often sad at getting home later, or missing time with you to study or write, I hope you see that hard work pays off. Chase every goal and never give up. Being your mom is my proudest achievement, and nothing can ever bring me more joy.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGMENTS .....	iii
DEDICATION .....	iv
LIST OF TABLES .....	viii
LIST OF ILLUSTRATIONS .....	ix
CHAPTER I –Introduction .....	1
CHAPTER II – Review of Literature.....	8
Rubrics in Education.....	8
Reliability and Validity .....	12
Rubrics in the Music Classroom .....	15
Rubrics in Group Ensemble Performance Assessment.....	19
Limitations of Rubrics .....	22
Rubric Development .....	24
CHAPTER III – Methodology .....	26
Rubric Development .....	26
Developing the Rubric .....	30
Proposed Usefulness of the Rubric .....	31
Limitations of the Rubric .....	32
Current Study .....	35

CHAPTER IV – Results.....	37
Quantitative Results .....	37
Qualitative Results .....	46
Bias in choral adjudication.....	48
Score sheets in show choir adjudication. ....	51
Usability of the Tyson Rubric.....	54
Benefits and limitations of the Tyson Rubric .....	55
Caption scoring. ....	58
Closing remarks. ....	62
CHAPTER V – Discussion.....	64
Implications.....	67
Limitations .....	77
Suggestions for Future Research .....	79
Conclusion .....	82
APPENDIX A – Tyson Rubric .....	83
APPENDIX B – Adjudicator Interview Script.....	86
APPENDIX C –Interview Transcripts .....	87
APPENDIX D –ISSMA Rubric .....	138
APPENDIX E – IRB Approval Letter .....	140



REFERENCES .....	141
------------------	-----

LIST OF TABLES

Table 1 Competition Demographics ..... 36

Table 2 Competition and CI Results ..... 41

## LIST OF ILLUSTRATIONS

Figure 1. Vocal Performance Categories .....	27
Figure 2. Visual Performance Categories .....	28
Figure 3. Show Design Categories.....	28
Figure 4. Tyson Rubric – Ease of Use .....	43
Figure 5. Tyson Rubric – Clear Descriptions of Each Category .....	44
Figure 6. Tyson Rubric – Appropriately Weighted Points for Each Category .....	44
Figure 7. Appropriately Addresses the Dimensions of Show Choir Performance .....	45
Figure 8. Effective Tool for Evaluating Show Choir Performance .....	45

## CHAPTER I –Introduction

Show choir ensembles, though often a polarizing issue among choral directors, have continued to thrive in the choral classroom. According to Weaver and Hart's (2011) historical look at show choirs, a clear start of the genre is difficult to define, rather, show choir appears to be an outgrowth of vaudeville and musical theatre. The authors define show choir as "a mash-up between a standard choir, a dance team, and a drama club, which competes with similar groups throughout the country" (Weaver & Hart, 2011, preface). In *Sweat, Tears, and Jazz Hands*, the history of show choirs is loosely traced back to names like "Fred Waring and His Pennsylvanians," a group that gained popularity in the 1920s and 1930s for performances of popular American music (Weaver & Hart, 2011). Waring's fame and influence continued to grow with the birth of the television variety show. Famous popular music ensembles, such as Perry Como, Andy Williams, Lawrence Welk, and Ed Sullivan, grew out of the Waring tradition and eventually led to the most recent TV show, "Glee" (Weaver & Hart, 2001). Clearly the influence of historical ensembles, radio, and television sensations has had an impact on choral music education.

Competitions have become synonymous with show choirs from the early days of the genre. Weaver and Hart (2011) cite the first show choir competition as the Northwest Swing Choir Festival started in 1968 by Hal Malcolm, a former choral director at Mt. Hood Community College in Oregon. The competition trend spread across the nation as more show choirs formed. Both show choir and competitive choral competition continue to spark debate among music educators.

Millard (2014) argues that competitions are external motivating factors, helping teachers to sustain and increase focus in the subject area of choral music. Damon Brown, a nationally recognized show choir choreographer, advocates for competitive show choirs by saying “competition, for good or for bad, promotes excellence in our art. Many folks think competition is unhealthy, yet the quality of competitive groups versus noncompetitive groups is vastly different, with only a few notable exceptions” (Weaver & Hart, 2011, p. 39). Austin’s (1988) study on the effects of competition for elementary band students echoed Brown’s assertion, finding that students in rated contest groups performed better than students who were a part of a comments-only group (p. 102).

However, some researchers and music educators find competition to be part of a token economy that promotes a love for competition, rather than a love for the process of music education (Austin, 1988; Millard, 2014). Despite Austin’s findings, a dissenting view was presented in an opinion article for the *Music Educators Journal*, in which Austin stated that “competitive goal structures tend to promote an egotistic type of motivation whereby children focus on social-comparison information and disregard instructional feedback addressing the actual quality of their effort or performance” (Austin, 1998, p. 23-24). The competition debate has not diminished the growth or existence of competitive choral ensembles, in particular, show choir, which Latimer attributes to parents, administrators, and community members who feel that success in evaluation (competition) is closely related to the perceptions of success of a music program (Latimer, 2013).

According to the most recent and complete set of data tracking show choirs in the United States, numbers indicate 1,525 high school and 472 middle school show choirs in

America (“Show Choir Community”, 2002-2020). Most of these programs are in midwestern states, though nearly every state has at least one choral program that includes a show choir.

In Mississippi, show choir continues to be a prominent feature of many choral programs. Showchoir.com lists 75 high school and middle school show choirs in Mississippi (2002-2020). Nearly all of these show choirs are competitive in nature, traveling to compete in local, regional, and national competitions. The first show choir contest in Mississippi was founded in 1987 by Dr. Mark Malone, the Choral Director at Pearl River Community College. The inaugural competition hosted five local groups on the Pearl River campus. The contest continues to be held each year and has grown considerably in size. Seven additional contests are now held in the state by both private and public high schools.

Showchoir.com is the most up-to-date database for national show choir competitions, touting 469 show choir competitions for middle and high schools each year (2002-2020). Competitions begin in early January and continue through April. Weaver and Hart (2011) estimate by the 1990s, most show choirs attended at least one competition a year, yet today many groups compete six or seven times per competition season. Contests are open to any group who pays the registration fee and chooses to compete.

Some contests are governed by a state organization, much like athletics; however, a large portion of these contests are independently run by the sponsoring school or group. Independent show choir contests bring in huge revenue and are often the largest

fundraiser for the group each year. In addition, national competitions like “Fame”, “Showstoppers”, and “Heart of America” host groups at an array of sites each year.

Ensembles are classified according to age/grade (junior high/high school), gender, and size category, which depends upon the number of singers and dancers on the stage at one time. Groups compete within these categories and awards are given to the highest scoring ensembles. Many contests extend the competition day to include both a preliminary and finals competition. The top scoring groups throughout the entire day of competition are invited to return the same evening for a chance to compete again for an overall championship title. Caption awards are also given for best vocal performance, best visual performance, and best show design.

Despite the proliferation of competitions, many states do not have strategies in place to provide valid and reliable adjudication measures, nor does any sort of national standard exist. While most traditional choral festivals and competitions fall under some type of state governance, show choir competitions are usually independent in nature. Thus, rules, regulations, and scoring can vary widely between even state competitions. In fact, Weaver and Hart (2011) go so far as to define adjudication sheets as “a nonstandardized judge’s checklist used to determine the strength and quality of a group. . .” (p. 161). The lack of any sort of standardization in adjudication can certainly devalue the possible merits of competition and negate the educational value of receiving feedback. Therefore, a reliable adjudication system that employs research-based data collection within an appropriate competitive and pedagogic framework might strengthen the validity and educational merits of show choir competitions.

From week to week, show choirs are evaluated on different score sheets with varying point systems and categories, few of which attempt to qualify scores in an objective manner. Most score sheets involve subcategories worth ten points each. While adjudicators may strive to be as consistent as possible, difficulty arises when many score sheets fail to qualify how points are earned or justified. Many scoring systems use a 10-point scale, with each category weighted equally, although this may not be the most valid and reliable form of adjudication. Wide gaps in scores between judges could suggest a lack of appropriate expertise, but more likely indicate the absence of common standard. Though the nature of judging a musical performance is subjective, adjudicators must ensure that scores and comments are consistent and as unbiased as possible.

The educational value of results and feedback should be kept at the forefront of discussions regarding improved adjudication methods. Appropriate instruction and feedback for improvement is difficult when teachers receive widely varying scores between adjudicators. A more reliable and less subjective scoring process could ensure that students receive valuable and consistent feedback on how to improve, even if the group does not receive a trophy. Noted arranger of pop music, Kirby Shaw, stated, "We must keep asking ourselves why we are teaching and what we want the results to be. Enabling students to understand and perform the elements of music and movement....is a prize that can never be summed up by a trophy" (cited in Thomas, 2005). Keeping education at the center of adjudication makes the competitive experience meaningful regardless of the final placement.

The theory of reliable and education feedback has been applied and incorporated into many choral classrooms as a part of individual assessment and into the performance



rubrics of many state choral and instrumental ensemble festivals, but has not yet affected how show choirs are adjudicated (Latimer, Bergee, & Cohen, 2010; Millard, 2014; Norris & Borst, 2007; Rogers, 1983). The purpose of the current study is to contribute to the limited body of literature regarding competitive show choir by introducing a previously created weighted rubric, the Tyson Rubric, to nine show choir competitions and answering the following research questions:

1. Is the Tyson Rubric a statistically reliable instrument for scoring show choir ensembles?
2. What are the user perceptions of the Tyson Rubric regarding ease of use, categories and construct definitions, and the perceived educational value of the rubric as an instrument for feedback?

Following the competitions, the rubric was tested for statistical reliability using an intraclass correlation coefficient. Additionally, the researcher surveyed and interviewed adjudicators from the nine competitions to determine user perceptions of the rubric, as well as emerging themes regarding issues in show choir adjudication.

Previous studies have concluded that rubrics are a reliable form of assessment for other large music ensembles, such as choirs, bands, and orchestras, thus, the Tyson Rubric is also expected to be a reliable scoring system. (Brakel, 2006; Hash, 2012; King & Burnsed, 2009; Latimer, 2007; Latimer, 2013). In addition to statistical analysis, the researcher used a survey, along with personal interviews, to determine user perceptions of the rubric in regards to the ease of use and appropriateness of category weights and descriptors.

At the time of the current study, no empirical research exclusive to show choir exists, though the genre is a part of choral programs across the nation. Theories of best practices, specifically regarding show choir and show choir competitions, currently inhabit opinion articles without supporting research. Though the research questions apply specifically to the Tyson Rubric, perhaps the most important contribution of the study will be to instigate further research pertaining to show choir.

## CHAPTER II – Review of Literature

Evaluation and assessment are key components in education, particularly as a critical element in improving learning through feedback (Eisner, 2007; Gipps, 2011). Appropriate educational assessment allows educators to diagnose student learning and respond with instruction based on that diagnosis, in a continuous loop of feedback and learning. The cycle of instruction, learning, and assessment is known as formative assessment. Sadler (1998) defines formative assessment as an “assessment that is specifically intended to provide feedback on performance to improve and accelerate learning” (p. 77).

Formative assessment is particularly valuable in the music classroom, both for individual and ensemble feedback (Asmus, 1999; Barry, 2009; Latimer, 2013; Murphy, 2007). Likewise, the use of formative assessment in music competition is integral to ensemble success, particularly since many music educators are evaluated by administrators and parents based on the outcomes of such competitions (Boyle & Radocy, 1987; Burnsed & Sochinski, 1983; Forbes, 1994). If choral programs and directors are to be evaluated on competitive assessments, then scoring instruments must be detailed, educational, and reliable (Barry, 2009; Forbes, 1994; Hash, 2012). Discussions regarding formative assessments often lead to the use of rubrics as valuable tools for such feedback.

### Rubrics in Education

The foremost accepted definition of a rubric can be attributed to Heidi Andrade (2000), who defines a rubric as a document that “describes varying levels of quality, from excellent to poor, for a specific assignment” (p. 13). Rubrics are valued by educators for

ease of use, as well as the ability to make grading more time efficient without losing the quality of feedback (Andrade, 2000; Brodie & Gibbings, 2009; Carlin et al., 2011; Wilson, 2006).

The effectiveness of rubrics for grading purposes is commonly accepted in the academic community (Arter, 2000; Hafner & Hafner, 2003; Shipman et al., 2012), though Andrade et al. (2008) suggest that rubrics can be beneficial as teaching tools. Rubrics provide clarity for how grades are determined and have resulted in positive feedback from students regarding increased understanding of the assignment due to familiarity with the rubric being used (Andrade, 2001; Lee & Lee, 2009; McKenna, 2007). Andrade, Du, and Wang (2008) found that students who used a rubric to self-evaluate their first draft had a higher quality of writing.

Rubrics provide accountability and serve as a part of the instructional process, rather than simply a tool for evaluation (Andrade, 2000; Johnson et al., 2019; Logan & Mountain, 2018; Sparrow, 2004). Rubrics have the potential to provide very clear, concise feedback justifying a given score. McKenna (2007) praises rubrics as tools to explicitly articulate task criteria and grading norms, while forming “a critical link between assessment and learning by foregrounding the purpose of the task and how the educator defines quality” (p. 22). Students who are evaluated summatively only receive feedback at the completion of an assignment, which does not allow the possibility to adjust and learn throughout the process (McKenna, 2007).

Rubrics can aid in judging progress, allowing the learner to improve before a final evaluation. Thus, rubrics can link assessment and instruction in a way that improves

clarity of expectations and often results in a higher quality end result (Garipey et al., 2016; Herro et al., 2017; Hung, 2013; Tractenberg et al., 2010).

Research on the topic of instructional rubrics as tools for feedback includes nearly every school subject from mathematics to writing. Sundeen (2014) found that when rubrics were used as a part of the instructional process, rather than solely a grading tool, the quality of writing for high school students was improved (p. 84). Similar studies in which the researchers created a criteria-specific rubric showed results suggesting improved writing quality across the curriculum (Sundeen, 2014; Timmerman et al., 2011).

Rubrics are widely used as tools for performance assessment. Some studies indicate increased reliability in scoring through the use of rubrics, particularly if combined with rater training, though in and of themselves rubrics are not necessarily more valid than other performance assessment measures (Jonsson & Svingby, 2007; Silvestri & Oescher, 2006). Rubrics must be combined with judges who are familiar with the evaluation and trained in order to be both reliable and valid tools for assessment (East, 2009; Rezaei & Lovorn, 2010; Stellmack et al., 2009). Stemler (2004) asserts that rubrics have the potential to be useful by imposing a certain level of objectivity in a subjective performance assessment and states that:

interrater reliability must be demonstrated anew for each new study, even if the study is using a scoring rubric or instrument that has been shown to have high interrater reliability in the past. Interrater reliability refers to the level of agreement between a particular set of judges on a particular instrument at a particular time. Thus, interrater reliability is a property of the testing situation, and not of the instrument itself. (p. 1)

Acknowledging the fact that rater training increases the effectiveness of rubrics as a reliable scoring method is an important component of the research at hand.

Rubrics also have an impact on special learner populations, both in behavioral and instructional outcomes (Lee & Lee, 2009; Stanford & Reeves, 2005). Lee and Lee's study (2009) examines the benefits of instructional rubrics on the classroom engagement behaviors and learning outcomes of a population with "mild mental retardation" (p. 396). The rubric for this study included standard objectives, as well as modified objectives for the students with mild retardation (IEPs). Results of the study indicated that class engagement behaviors improved among those with mild mental retardation with the implementation of the rubric, and the behavior did not regress once the rubric was discontinued. The authors suggest that the rubric creates awareness of targeted lesson objectives and increased self-awareness of student performance. Ultimately, students both with and without special needs were shown to benefit from the instructional rubric in the inclusive setting (Lee & Lee, 2009).

In a similar study, Stanford and Reeves (2005) found that assessment techniques inform instructional strategies, particularly when dealing with students who have learning disabilities. The authors propose that assessment rubrics may be helpful in improving both instruction and assessment for this population. Rubrics, therefore, serve as a decision-making tool for instruction providing a way to learn through assessment. Stanford and Reeves (2005) refer to this as a "continuous feedback loop," which aids the learner in determining the exact criteria of assessment (p.20).

According to a meta-analysis by Jonsson and Svingby (2007), studies that examine teacher and student attitudes towards rubrics as assessment tools are

overwhelmingly positive. Specifically, rubrics are valued for making expectations explicit. Interestingly, the positive attitudes continue to exist regardless of actual outcomes, meaning that students believe learning is better with rubrics, even if the assessment does not indicate increased learning (Jonsson & Svingby, 2007).

### Reliability and Validity

Validity, according to Kural's (2018) definition, is "a general term which is related to questions about what the test is actually assessing" (p. 343). In other words, validity shows that the test, assessment, rubric, et cetera, measures the intended outcomes. If an evaluation intends to measure a student's sight-reading ability, but actually measures the student's tone quality, then the evaluation is not valid. Kural (2018) defines reliability as "an estimate of the consistency" of an evaluative tool or assessment (p. 344). A reliable evaluation is one that produces the same scores across different time domains and with varying examiners or adjudicators.

Rubrics are praised for consistency in reliability and validity. Jonsson and Svingby (2007) found that rubrics, in and of themselves, are not necessarily more reliable than other means of performance assessment; however, they are more likely to produce reliable assessment than other open-ended measures. For example, research indicates that tests including multiple choice or true/false questions produce more reliable results than tests with open-ended discussion or essay questions (Kural, 2018). The authors also assert that validity is the more critical issue, though validity is more difficult to determine than the overall validity of rubrics, as previous studies tend to focus only on one aspect of validity (i.e. construct, criterion, and content) rather than validity in holistic terms (Jonsson & Svingby, 2007).

Silvestri and Oescher (2006) address concerns regarding subjectivity and lack of reliability in the assessment process, stating the rubrics objectify the scoring process through clearly delineating criteria. The authors' study compares the reliability of scores using both a rubric and an alternative performance-based assessment for student knowledge of health-related issues. The researchers developed a 5-point scale scoring rubric to assess student performance, then rated four sample papers independently and compared those scores to attain a "true" score. Sixteen pre-service teachers then rated the sample papers with and without the rubric. The results show that scores were inflated when the pre-service teachers did not use the rubric, while the scores when using the rubric were more comparable to the "true" score of the researchers.

The authors conclude that alternative assessments (those without a rubric) allow for subjectivity and thus reduce the reliability of scores. The sample also reported that adjudicators scored more quickly, felt less subjective, and their scores were more consistent when using the rubric (Silvestri & Oescher, 2006).

Rezaei and Lovorn's (2010) experimental study investigates the validity and reliability of rubrics used to score social science writing prompts. Rubric reliability was determined using standard error of measurement, as previous studies indicate this is most reliable for criterion-based assessment. Participants were asked to score written essays on a social science topic. The essays were scored twice by each participant, once using the rubric and once without using the rubric. The rubric was weighted, with 75 points dedicated to content and understanding, 15 points dedicated to proper citations, and 10 point dedicated to mechanics. The non-rubric grades were determined using a global 100-point scale. In this particular study, the use of a rubric did not increase reliability by



lessening score variance. Rather, scores determined by the use of a rubric increased in variability. Furthermore, the rubric affected the scorers grading focus. Without the rubric, raters focused more on content than mechanics. When using the rubric, raters deducted more points for mechanical errors, seeming to overlook content. The authors also interpret results to indicate that familiarity with scoring rubrics in general is not enough to ensure they are used effectively. Instead, raters should be oriented to specific rubrics and assignments/assessments before scoring. Rezaie and Lovorn (2010) state that “ideally, the feedback given by employment of a rubric is better than the assignment of a simple letter grade; however, untrained users of rubrics may simply use it to justify their biased assessment” (p. 10).

Stellmack, et al., (2009) evaluated a rubric used to grade APA-style introductions in an undergraduate population and found that, while successful as an instructional tool, the rubric did not eliminate the inherent subjectivity of grading writing assignments. The validity and reliability of the rubric was comparable to other similar rubrics; however, the authors suggest that “merely using an explicit, carefully developed rubric does not guarantee high reliability” (Stellmack et al., 2009, p.1). This echoes much of the existing literature on rubrics as instructional tools in that, while rubrics may help guide, organize, and instruct, but do not eliminate all sources of subjectivity and bias (Gerritson, 2013; Milanowski, 2017; Wheat Sr. & Bartlett, 1982;). Unless grades are assigned in a blind rating process, raters may be influenced by a variety of factors.

According to Stemler (2004), “judges are most often used when behaviors of interest cannot be objectively scored in a simple right/wrong sense, but instead require some rating of the degree to which observed behaviors represent particular levels of a

construct of interest” (p.1). In such situations, rubrics are often used to reduce subjectivity. However, Stemler notes that it is impossible to eliminate all subjectivity, due to a judge’s personal interpretation of a given construct within a rubric. The researcher further suggests that interrater reliability is one of the most “important concepts in educational and psychological measurement,” and “without demonstrating that two independent judges can be reliably trained to rate a particular behavior, our hope for achieving objective measurement of behavioral phenomena is diminished” (Stemler, 2004, p. 9). With this statement, Stemler alleges that the absence of interrater reliability ensures that objective scoring or assessment is unattainable. Stemler (2004) argues that comprehensive references to interrater reliability without proper contextualization are misleading and do not fully represent the range of true reliability. Stemler defines interrater reliability as “the level of agreement between a particular set of judges on a particular instrument at a particular time. Thus, interrater reliability is a property of the testing situation, and not of the instrument itself” (Stemler, 2004, p. 1).

### Rubrics in the Music Classroom

Though assessment is a vital element of the music classroom, it presents a host of challenges and uses, though music educators seem to agree that it is a vital element of the music classroom (Asmus, 1999; Barry 2009; Latimer, 2013; Latimer et al., 2010). Latimer (2013) states that evaluating musical performances differs from other mediums: Unlike other content areas (for example, math, reading, social studies, and so forth), which generally focus on assessment of written examples of student work (usually sometime after they are written), music assessments frequently involve evaluations of

live performances in real time, that is to say, as they are happening and immediately after they are over (p. 81).

Asmus (1999) states, “It should be understood that assessment is not an add-on to instruction. Rather, it is an integral part of the instructional process, and it can inform both the teacher and the learner” (p. 19). Similarly, Barry (2009) discusses the longstanding tradition of music performance evaluation as an integral part of music education. Latimer, et al. (2010) agree that “music rubrics have been employed widely in academic and performance contexts, often with the goal of providing more specific information to student musicians and better evidence of student progress toward benchmarks and standards than non-rubric assessment forms” (p. 170).

While artistic evaluation is inherently subjective, Barry argues that “demonstrated mastery of certain technical standards is expected of students in the arts,” resulting in a greater need for integrous evaluation, largely due to political and social pressure (p. 246). For example, the components of *No Child Left Behind* focus largely on high stakes evaluation, rewarding schools that perform well. This rationale has infiltrated higher education as well, with greater emphasis on “accountability and transparency” (Barry, 2009, p. 247). Barry states that “evaluating music performance. . . has always presented challenges with respect to balancing the subjective, personal nature of artistic performance with the need to maintain some degree of consistency and objectivity” (p. 249).

Barry also highlights the difficulty in evaluating musical performance. Some studies have revealed that faculty evaluations of student performance may be highly unreliable and even biased on the basis of influences such as the time of day,

performance order, and even the performer's attractiveness, stage behavior, and dress. Reliability tends to be quite high, however, when carefully-developed tools such as criterion-specific rating scales and rubrics are used. (p. 250)

The music classroom makes particularly effective use of rubrics as a means of assessment (Asmus, 1999; Barry, 2009; Murphy, 2007). Aside from the general music/music appreciation course, most instruction in music evaluates based on skills in a performance environment, both individually and collectively. Ciorba and Smith (2009) developed a rubric for applied jury assessment at a university with a large number of music majors. The rubric was used to evaluate jury performances across multiple mediums. Results yielded a significantly higher level of both internal and interrater reliability, though such evaluation presents a conundrum for evaluators.

Students learn skills such as sight-reading, aural training, phrasing, vocal flexibility, range development, dynamics, blend, balance and breath management for participation in ensembles, which are all necessary and inherent components of a performance-based music education. Eisner (2007) comments that “one cannot be adequately engaged in teaching without also being engaged in an evaluative process” (p. 423). Eisner (2007) further asserts that evaluation and assessment are an inherent part of the teaching and learning process, stating that “assessment and evaluation processes are educational tools; they can be used to inform policy and contribute to improved educational practice” (p. 424).

Latimer (2013) describes the assessment process in regards to musical performances as “broken into component parts, or musical dimensions. . . which are often scored, totaled, and then presented in aggregate in the form of a global assessment: a

numeric score, grade, or rating” (p. 82). Rubrics have gained popularity among music educators to accurately assess performance-based skills and assign such ratings with confidence in the justifications corroborating the score (Stanley & Brooker, 2002; Wesolowski, 2018).

Wesolowski (2012) sums up what is inherently unique and problematic about adjudicating any musical performance, stating “Music...is a discipline that embraces expressive decisions and divergence of response” (p. 37). Essentially, authentic assessment must be objectively subjective. That is to say, while one cannot dismiss the personal philosophies and experiences that influence individual beliefs about musical performance ideals, one must strive for consistency and fairness in our evaluations. Personal preference becomes a challenge to overcome when assessing students and ensembles.

Eisner (2007) echoes these sentiments, stating that the difficulty lies in the ability to “see the qualities that constitute virtue in an art form and to have the articulateness to provide reasons for choices made by the assessor pertaining to quality of work that has been created.” (p. 426). By using the specificity of a well-designed rubric, some level of objectivity can be applied to assessing musical skills. Though Wesolowski’s (2012) research indicates that some judges may feel compartmentalized by criteria-specific assessments, those same judges concede that these types of assessments help to increase focus and communication of feedback. Similar to research results in general education, music research points to the value of assessor training to ensure the reliability and validity of criteria-based scoring (Stegman, 2009; Wesolowski, 2012). Some states, such as Michigan and Indiana, have greatly revised adjudication measures to include valuable

training. According to Stegman (2009), Michigan adjudicators must attend an all-day clinic to be certified to judge a state choral festival. Additionally, Mississippi now requires that all adjudicators for choral festivals must attend a mandatory training workshop every five years in order to remain on the approved adjudicator list (Mississippi Music Educators Association, 2020, para. 2). Similarly, Indiana State School Music Association requires that all potential adjudicators complete an online training course (Indiana State School Music Association, n.d.).

### Rubrics in Group Ensemble Performance Assessment

Wesolowski (2012) defines a rubric as a “form of criteria-specific performance scale.... [that] divides a task into constituent parts and offers detailed descriptions of the performance levels for each part” (p. 37). The addition of specific qualifiers for each rating option of every category gives the adjudicator the opportunity to make rating decisions that align with standards, rather than assigning an arbitrary number. Norris and Borst (2007) agree with this opinion, stating that using rubrics keeps judges in line with specific criteria. Teachers and students alike tend to prefer specificity in feedback, as each is able to focus on areas of proficiency, and those in which improvement is needed. The critical feedback loop (Johnson, 1978) occurs when information from an assessment results in improved instruction and better performance, a theory that can be practiced in choral ensembles. Asmus (1999) suggests that using a rubric to evaluate ensembles in rehearsals prior to festival performances may better prepare the ensemble for the rated performance, thus utilizing the feedback loop to continually improve.

In a 2010 article, Latimer et al. (2010) describe the results of a study comparing a weighted performance rubric in Kansas large ensemble festivals with unweighted

assessment tools that were used previously. A KMEA committee, consisting of choir, band, and orchestra directors, as well as KMEA board members, developed the weighted rubric after conversations with directors indicated that such a rubric would be a more descriptive teaching tool than previous score sheets. The resulting rubric included the same nine categories for band, choir, and orchestra, with the exception of the technique category for band, which was replaced with diction for choir. The rubric contained the following dimensions with corresponding points: Tone (15), Intonation (15), Expression (15), Technique or Diction (10), Rhythm (5), Note Accuracy (5), Balance (5), Blend (5), and Other (5) (Latimer et al., 2010, p. 171).

The study found that many adjudicators believed scores were better justified when using the rubric and directors valued the increased amount of specific feedback provided. Results of this study indicate a moderately high level of reliability of the KHSAA rubric. Interestingly, like Norris and Borst (2007), the dimension of rhythm was found to be less reliable than other categories. Results also indicate that scores become less reliable as each category descends in weight. In other words, the dimensions on this particular rubric are ordered from highest to lowest score (the rubric is weighted, meaning that not every dimension is worth the same amount of points). The authors attribute the decrease in reliability to score restriction, and also suggest the possibility that ordering the dimensions by weight may influence scores.

King and Burnsed (2009) examined the scoring at a Virginian Band and Orchestra festival using a non-rubric score sheet that had high reliability for overall scores. The ensembles were adjudicated on five dimensions (captions): Quality of Sound, Technique, Musicianship, Technique and Ensemble, and Design. However, the caption awards did

not hold the same reliability as the overall scores. The authors suggest that adjudicators may be giving overall ratings without consideration to the individual captions or dimensions, giving concern to the validity of the scoring procedure. The researchers suggest that a criteria-specific rating scale, more descriptive like a rubric, may improve the validity of scoring by measuring distinct performance elements more accurately.

Latimer, Bergee, and Cohen's (2010) research is similar to the current research regarding the Tyson Rubric, in that it examines both the function of a rubric as a reliable scoring method and the pedagogical implications as perceived by adjudicators and directors. Latimer's (2013) longitudinal study examining the validity and reliability of a large group performance assessment rubric used in a midwestern state acknowledges that criteria-specific rubrics do not necessarily hold greater reliability or validity than traditional or Likert-type scoring sheets. However, the study proposes that rubrics provide more detailed feedback that hold greater educational value than the traditional score sheet that lacks descriptive information or requires the adjudicator to simply circle an overall numerical score without providing justification. This coincides with a general educational trend towards "learner-centered systems" of evaluation (Latimer, 2013, p.86). Furthermore, Latimer suggests that rubrics may merely be perceived to be a better assessment by judges and directors. Wesolowski (2012) concurs with Latimer's opinion of the value of rubrics for educational feedback, rather than solely for scoring or ranking, and argues that rubrics are able to more richly and specifically assess music students, adding to individual growth as musicians.



## Limitations of Rubrics

Rubrics, though shown to be reliable and valid in assessment, are not without limitations. The very nature of performance adjudication is subjective, which may not be without merit. Adjudicators are chosen for their particular expertise and body of experience which aids in judgement. Murphy (2007) cites previous authors who argue that subjectivity is important in the evaluation of musical performance, although bias can occur and reduce the reliability of assessment measures. Contest and festival directors must balance subjectivity and mitigate bias. While utilizing a criteria-based assessment is one way to overcome bias and improve reliability, Murphy (2007) states that “criteria can impose limitations on examiners that undermine the assessment process, emphasize a narrow view of music performance characteristics and even interfere with holistic judgment” (p. 372).

Overcoming bias presents a challenge for teachers and adjudicators because it is also important to recognize the individuality of both the adjudicator and the learners or performers being assessed. Wilson (2006) asserts that many rubrics fail to consider individual values and aesthetics of both the learner and the assessor. Additionally, rubrics fail to account for the “non-standard” learner, such as the individual limitations of those with learning disabilities and, according to Wilson (2006), failure to take such issues into account limits the educational properties of rubrics. The author goes so far as to compare the use of rubrics with high stakes testing, saying that a rubric “doesn’t reflect any other value we might hold about assessment, including . . . reflection of a student’s efforts” (Wilson, 2006, p. 28). This type of thinking towards assessments creates a mindset that is increasingly focused on finding flaws in order to justify a score. Rubrics often restrict

educators into judgment decisions that exclude special circumstances and overlook outside factors, such as effort, improvement over time, and work ethic.

Though rubrics may improve reliability in scoring, other factors may still inhibit unbiased adjudication. Wesolowski, Wind, and Engelhard (2015) suggest several types of rater errors: clashing standards, clashing values, fatigue, halo effect, length, personality clash, repetition factor, and sympathy scores, which are independent of the musical evaluation, but have been shown to affect scores. Clashing standards, values, and personalities refers to the individual biases of adjudicators, which may oppose the standards and values reflected by other adjudicators, ensemble directors, or the scoring rubric itself. The rater errors of fatigue and length may also affect scores. As contests often last for 12 or more hours a day, or even over an entire weekend, raters become physically and mentally tired and may be less likely to score accurately.

The halo effect may also play a role in evaluation. Boyle and Radocy (1987) define the halo effect as a “positive or negative bias because of what the reader knows about the writer”, or in this case, what an adjudicator knows about a director or ensemble (p. 123). According to Forbes (1994), the halo effect may be mitigated by choosing adjudicators from outside the region where the contest is held, choosing adjudicators who have no prior knowledge of or experience with the competing ensembles, or using pre-recorded performances for blind adjudication. Forbes also advises against supplying adjudicators with informational sheets about the ensembles.

Repetition of repertoire may negatively affect ensemble scores. As adjudicators hear the same song choice multiple times throughout a competition, or even across different competitions, they become desensitized to each individual performance. It

becomes increasingly difficult to appropriately adjudicate and differentiate between performances of the same song. Finally, sympathy scores are one of the most common types of rater errors. In trying to avoid scoring in an overly harsh manner, adjudicators may boost scores of lower performing groups, often abandoning the rubric. Cassidy and Sims' (1991) research regarding scoring of a handicapped youth choir indicated that both peer and adult evaluations were affected positively with knowledge of the ensemble's label. The ensemble scored more highly with evaluators who knew it was a handicapped choir, while evaluators who did not know the group's label scored them lower, suggesting that sympathy scores did play a role in adjudication.

In Wesolowski's 2015 study, results may be interpreted to suggest that raters scores fluctuate depending on the school level of the performing ensemble (middle school, high school, etc.). Non-musical factors may play a role in scoring, despite the best efforts of score sheet standardization and criterion-specific rubrics.

### Rubric Development

The body of literature dedicated to developing rubrics for the assessment of ensemble musical performances is limited. Barry (2009) and Wesolowski (2012) provide similar approaches to creating rubrics for performance assessment. Barry (2009) suggests that the first step is to develop the list of dimensions of the performance that will be evaluated collaboratively with those who will be assessing the performance. Wesolowski (2012) adds that educational outcomes or objectives should be included in the list of performance dimensions. Next, rubric creators must further specify the criteria for each dimension, differentiating between levels of proficiency in language that is clear. Though both authors encourage revision throughout the development process, Barry

suggests a pilot test to ensure that the instrument is valid, reliable, and easy to use before employing a rubric in an actual situation.

While a criteria-specific performance scale has been applied and incorporated into many choral classrooms as a part of individual assessment and into the performance rubrics of many state choral festivals, the rubric has not yet affected how show choirs are adjudicated. In light of the absence of any sort of standard adjudication form for show choir contests and given the amount of research suggesting the value of criteria-based assessment in the form of a rubric, a pilot study was conducted in the fall of 2016. The end goal of the pilot study was to develop and put into practice a rubric for use in show choir contests, as a means of creating a more valid and reliable standard for adjudication.

## CHAPTER III – Methodology

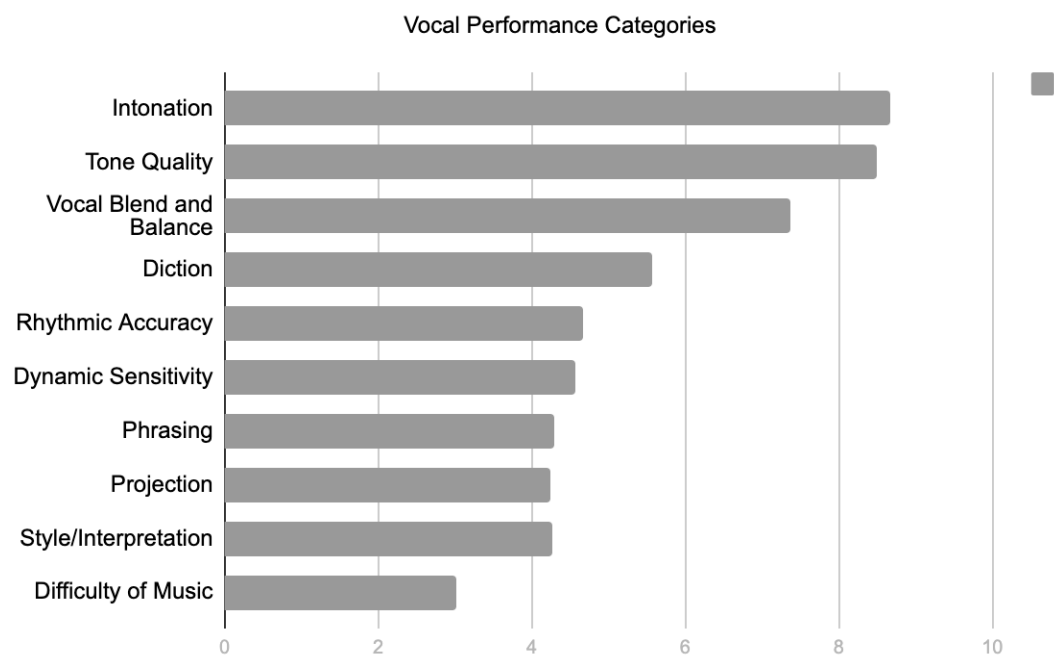
### Rubric Development

In a pilot study, the researcher developed a weighted rubric for scoring show choir contests. A weighted rubric assigns points to subcategories, with some receiving more than others, based on perceived value or importance. A typical score sheet for show choir contests includes three major sections: vocal performance, visual performance, and show design (also labeled as overall effect for some score sheets). Each of the sections includes various subcategories, each usually worth ten points, that may differ from contest to contest. The vocal performance section usually includes more subcategories than the other two, giving more total points to the area of vocal performance.

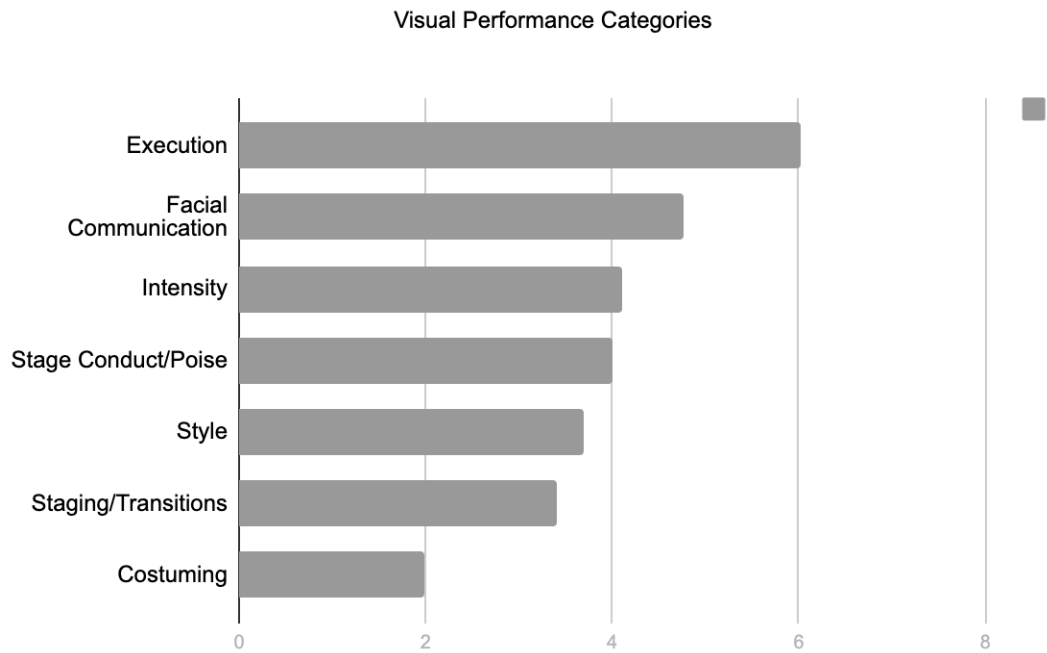
In order to create a weighted rubric, the researcher first had to determine what categories should receive the most weight. A survey was developed that determined participant demographic information and then asked raters to rank the typical subcategories of each larger section in order of importance, most to least. The only exception was the Overall Effect section, in which respondents chose the most important of the three subcategories rather than ranking each in order of perceived importance.

The survey link was emailed to show choir directors, choreographers, and adjudicators within the Southeastern region of the United States. Links were also posted on the researcher's personal Facebook page, as well as the group pages, "Show Choir Directors and Choreographers" and "I'm A Choir Director." A total number of 74 people responded to the survey. Of these people, 75% were currently teachers in a K-12 music program, with 74% directors of a competitive show choir. The teachers were equally distributed between the South and Midwest regions of the U.S.

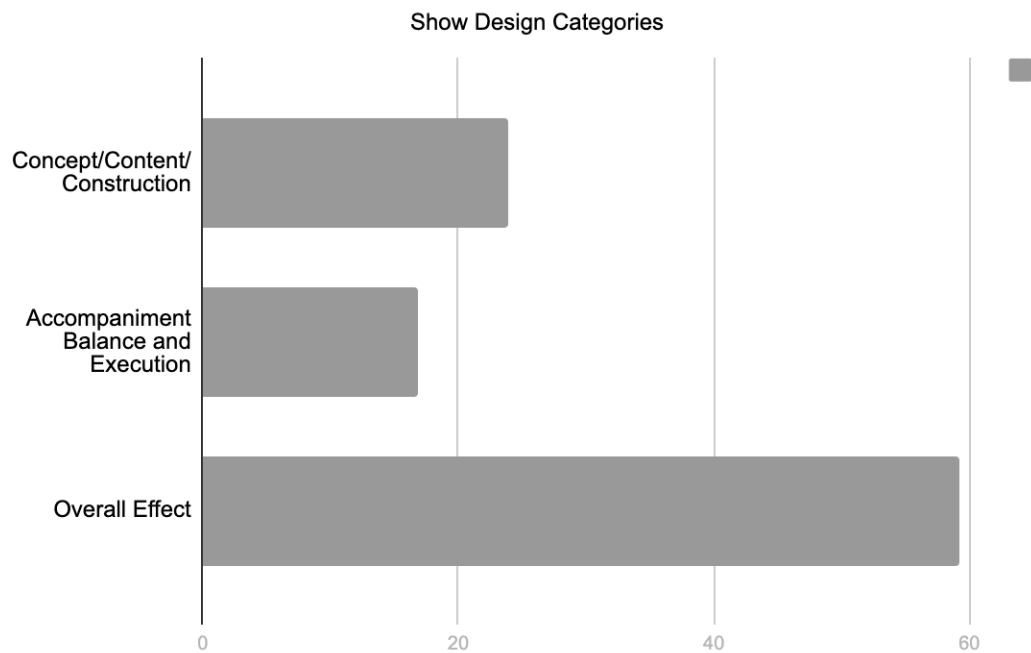
When asked about the level of satisfaction with scoring methods, 39% of teachers indicated that they are dissatisfied with scoring methods in the school's geographical region. Only slightly less than 6.85% of teachers indicated extreme satisfaction with scoring methods in the respective region. Teachers were asked to rank the categories from a typical score sheet in order of importance on a scale of 1-10. Figures 1, 2, and 3 display the results.



*Figure 1. Vocal Performance Categories*



*Figure 2. Visual Performance Categories*



*Figure 3. Show Design Categories*

\*Participants were asked to choose which category was most important.

In the Vocal Performance area, the sub-category of intonation received the highest rating (8.67%), followed by tone quality (8.49%), and vocal blend and balance (7.38%). The next highest category was diction (5.58%). The remaining categories of rhythmic accuracy, dynamic sensitivity, phrasing, projection, style/interpretation and difficulty of music, varied little in terms of perceived importance. However, difficulty of music (3.01%) was deemed least important by a relatively large margin.

The Visual Performance subcategories show a greater difference in score between each component. Execution received the highest rank (6.03%). Facial communication was regarded as the second most important aspect (4.77%), followed by intensity (4.11%), stage conduct/poise (4.01%), style (3.69%), and staging/transitions (3.41). Costuming received the lowest rank by a considerable percentage (1.99%). Teachers were given the opportunity to answer an open-ended question regarding the parts of the current scoring methods that are most problematic. Some general themes were revealed in the responses given. The most common source of discontent was the inconsistency of scoring rubrics between competitions and the absence of a more universal or standardized scoring method. In addition, teachers believe that the current methods do not alleviate adjudicator bias and subjectivity. The lack of justification for scores leads to much frustration for teachers, students, parents, and the general public in the audience. Another common theme was the qualification, experience, and training level of the adjudicators. Future research is needed to determine the extent to which adjudicator training affects scoring, as well as the amount and type of adjudicator training that is necessary.



## Developing the Rubric

Though the survey results certainly informed decisions made regarding category weights, the researcher's own experience as an adjudicator and director also contributed to the development of the rubric. Time was spent interviewing colleagues, both fellow directors and adjudicators, in-person and through email. The resulting rubric reflects the views of the creator, as well as those of the survey participants. For each of the categories (Vocal Performance, Visual Performance, and Show Design), subcategories were weighted based on perceived importance, as shown by the survey results. Those subcategories receiving the highest rank received the most point values in the rubric. For example, intonation and execution, each the highest-ranking subcategory in the larger categories of vocal and visual performance, received a value of ten points in the rubric. Middle scoring subcategories, like dynamic sensitivity, were given a value of seven points. The lowest scoring subcategories, such as difficulty of music and costuming, only received a value of five points. Thus, the resulting rubric is reflective of the survey responses.

Once the weights for each category were determined, operational definitions were added to the rubric and altered to reflect the levels of performance for each category. Categories receiving ten points were divided into proficiency levels labeled as Superior, Excellent, Good, Fair, and Needs Improvement. Categories receiving seven points were divided into levels labeled Excellent, Good, Fair, and Needs Improvement. Categories receiving five points were divided into levels labeled Excellent, Good, and Needs Improvement. Those reviewing the Tyson Rubric generally confirmed the validity of the

resulting adjudication form, although interesting discussions and arguments were made for and against the inclusion and weights given to some categories.

#### Proposed Usefulness of the Rubric

The show choir adjudication project began as a way to improve the adjudication procedures of show choir contests, yet the finished product has implications for use other than that of its original intent. As discussed earlier, competitive show choir can provide an effective environment for exploring educational outcomes. Students who participate in this particular arena have the potential to become not only outstanding musical performers, but also educated critics of musical performances. Evaluative tools are useful outside of the role of the adjudicator and are a part of NAFME National Standards for music education, as well as near the apex of Bloom's Taxonomy.

In trying to improve adjudication methods, the educational value of scores and feedback should be kept at the forefront of the discussion. A more reliable and less subjective scoring process can ensure that students receive valuable and consistent feedback on how to improve, even if a trophy is not awarded. Kirby Shaw (2005) stated that, "We must keep asking ourselves why we are teaching and what we want the result to be. Enabling students to understand and perform the elements of music and movement....is a prize that can never be summed up by a trophy" (p. 108). Keeping education at the center of adjudication makes the competitive experience meaningful regardless of the final placement.

A research-based rubric could also prove useful for individual assessment in the classroom. Choral students are often assigned grades based solely on attendance and participation, but more specific feedback to an evaluation of personal musical growth

would be more in line with evidence of mastery of national and state standards for music education. Wesolowski (2012) asserts that an analytic rubric allows the student, as well as the teacher and parents, to clearly see individual strengths and weaknesses.

A rubric can also be used as an educational tool both prior to and after a performance. Trying to explain tone quality, dynamic sensitivity, or intensity in choreography from the feedback given by 4-5 judges, each with a different opinion of what that means in context and scores often varying by several points, lacks pedagogical pragmatism. The definitions within the rubric can provide a context for classroom discussions beyond a numerical score. By educating students on the value and importance of specific criteria, each will grow as a musician and, in turn, have the opportunity to gain higher scores in competition. The debriefing period after a competition can be used to explain with better specificity where the group needs to improve and provide a roadmap for future rehearsals.

#### Limitations of the Rubric

While a research-based rubric is arguably a step in the right direction towards improving standardization and fairness in show choir adjudication, it does not resolve all of the issues with choral competition judging. Most of the survey respondents were from two regions of the United States, the South and the Midwest. The focus of show choir in Midwest and Southern regions is similar in nature, but for East and West Coasts, values, as well as, scoring methods may differ greatly. The Tyson Rubric may not be representative of the aspects of performance valued in all regions.

Many of the suggestions for other categories to be included or removed did not affect the construction of the rubric. For example, some survey respondents felt that

props and backdrops should be evaluated as well. Other respondents felt that staging/transitions should not be a part of the rubric at all. The most contested subcategories were “difficulty of music” and “intensity.” “Difficulty of music” is challenging, not in definition, but in retaining educational and philosophical values in that explanation. Furthermore, since judges are not provided with written notation, each makes judgments on the difficulty of music based only on what is being heard. Future research is needed to continue refining the performing aspects being assessed. In regards to the dimension of “intensity” in the section of visual performance, many remarked that while easy to see, “intensity” is difficult to define.

Categories, such as rhythmic accuracy and accompaniment, are also debated in regard to personal opinions and competition rules that cannot be addressed in the Tyson Rubric. The largest debate concerns accompaniment, and most specifically live accompaniment. Many respondents expressed a need for clarification and continuing discussion on how to appropriately score recorded accompaniment versus a live band. The rules for scoring bands vary widely across competitions. Many competitions have rules that only allow a certain number of adults to play in order for the group to win a “Best Band” caption award. Some teachers believe that band should not be scored at all; while others contend that a show choir performing with an accompaniment track should not receive the same score as those that are accompanied by a live band. Further discussion is needed to clarify these views and standardize the corresponding rules and scoring methods for this subcategory. Until more agreement is reached on accompaniment, changes to the rubric are unlikely to be beneficial.

Many survey respondents suggested that a main problem in the realm of competitive show choir is the lack of qualified and adequately trained judges. The hiring and training of judges is the responsibility of each contest producer. Opinions on adjudicator competency vary widely and would be a worthy subject for future research. As suggested by Wesolowski, Wind, and Engelhardt (2015), for a performance assessment rubric to be effective, it must be accompanied by a “combination of rater training, development of exemplars and anchors to accompany the scale, and a clear benchmarking and cut score system be developed” (p.165). Through research and experience, the researcher prescribes that adjudicators be provided with a brief orientation to the Tyson Rubric prior to the start of the competition, mainly because it is such a departure from previously used score sheets. Show choir contest producers may also want to email a copy to each judge a week before the contest to give adequate time to become familiar with the rubric. The issue of adjudicator training is one that the research plans to investigate based on results of the study. The final research product will include a prescription for adjudicator training materials.

The most obvious limitation to the rubric is that it is reflective of the researcher’s own personal philosophies. Though the survey certainly informed the creation of the rubric, a higher value was placed on some aspects above what the results indicated based on the experiences and values of the rubric creator. For instance, dynamic sensitivity is worth more than the survey suggested. Musicians have preferences that are reflective of personal beliefs and experiences with music, and these constructs played a large role in the creation of the Tyson Rubric. Author Mara Wilson (2006) discusses the negation of the assessor’s values as a problem with rubrics in general and argues that personal values

and experiences are what qualifies the assessor to grade or adjudicate. Balancing personal values with assessor objectivity will always be a limitation of any rubric.

While further research and discussion is needed to continue to improve adjudication measures in the realm of competitive show choir, enhanced assessment tools are a step in that direction. Rubrics provide specific feedback that allow for subjectivity, while providing reliable justification of objective numerical scores. Such specificity allows for greater discussion in the classroom and more distinction between strengths and weaknesses of a performance.

### Current Study

The researcher used a mixed method design to answer the research questions. A quantitative approach applied statistical analysis to determine if the Tyson Rubric is a reliable method for scoring show choir competitions. User perceptions of the Tyson Rubric were determined through a quantitative survey and qualitative interview process.

Raw score data was collected from the use of the Tyson Rubric at nine competitions for high school show choirs during the 2018-2019 competition season. Except for one contest in West Virginia, all of the competitions took place in Mississippi. The competitions use the Carmen Scoring System, a web-based application that allows scoring rubrics to be collected in the cloud. Carmen Scoring Systems input the Tyson Rubric, allowing judges to record scores online, and also tabulated the scores for each judge, as well as for the entire panel of adjudicators. Table 1 indicates the number of judges and performing groups at each contest.

Table 1  
*Competition Demographics*

Competition	Number of Performing Groups	Number of Judges
Jackson Academy Show Choir Invitational	14	3
Jackson Prep Show Choir Masters	15	4
South Jones Invitational	15	3
Mississippi Showchoir Contest	15	3
Oak Grove Magnolia Invitational	11	3
Petal Show Choir Invitational	22	3
Purvis Show Choir Invitational	12	3
Robert C. Byrd Vocalfest	9	3
West Jones Show Choir Invitational	18	4
Total	131	29

Table 1 *Competition Demographics*

Following the competitions, the raw data was statistically analyzed to determine the interrater reliability of the rubric using an intraclass correlation coefficient. Following each contest, the judges were emailed the link to a survey measuring the user perceptions of the rubric. In addition to demographic information inquiries, the survey included questions answered on a Likert scale ranging from strongly disagree to strongly agree. Adjudicators responded to questions regarding the ease of use of the Tyson Rubric, the categories and construct definitions included in the rubric, and the perceived educational value of the rubric as an instrument for feedback. Additionally, some judges were asked to participate in interviews via phone or live chat to provide further feedback.

## CHAPTER IV – Results

### Quantitative Results

The study of the Tyson Rubric used an intraclass correlation coefficient to determine interrater reliability, which is the degree of agreement among raters, or how consistent raters are relative to one another in measuring a quantitative outcome. In other words, it provides an assessment of the consistency and conformity of measurements made by different raters who rate diverse participants on the same outcome. Stemler (2004) defines interrater reliability as the “level of agreement between a particular set of judges on a particular instrument at a particular time” (p.1). One way to determine interrater reliability is through the use of consistency estimates. By Stemler’s definition, consistency estimates do not require judges to “share a common meaning of the rating scale as long as each judge is consistent in classifying the phenomenon according to his or her own definition of the scale” (2004, p. 3). Similarly, Shrout and Fleiss (1979) acknowledge the need of assessing reliability indices when taking human measurement error into account, declaring that the intraclass correlation coefficient (ICC) is one way to ensure reliability between judges. Intraclass correlations, according to the *Encyclopedia of Statistics in Behavioral Science*, “are used to assess the consistency between measures of the same class,” including “comparing paired data on the same measure and assessing the consistency between judges’ ratings of a set of objects” (Field, 2005).

Computing the ICC requires that certain assumptions about the data are met. First, the assumption of independent observations was met as each judge rated each performance/contestant only once. Furthermore, judges did not confer with each other



about personal ratings on performances. Second, the data was continuous as the sum scores of each judge's rating for each performance/contestant within competitions was used in the computation of ICC. Finally, visual inspection and calculation of the pseudo  $z$  were used to test the assumption of normality. This assumption was met as inspection of histograms showed distributions that appeared normal. Also, pseudo  $z$  scores for skewness and kurtosis were well within the absolute value of two standard deviations. Across the samples of ratings from the judged competitions, the pseudo  $z$  scores for skewness ranged from  $-.80$  to  $.80$ , while those for kurtosis ranged from  $-.93$  to  $.85$ . Therefore, normality was assumed.

For each competition, the researcher compiled the sum (total) scores reported by each judge for each ensemble, which resulted in nine different data sets representing total scores reported by each judge for all performances within a competition. The resulting data was used in the computation of the ICC, thus, nine ICCs were computed in total. The researcher then selected an analysis model to account for the possible error that can influence the relationship based on how raters and samples were chosen. The selected model assumed that the same raters were used to rate all ratees (within each competition) and the sample of raters were drawn from a population of possible raters. The Two-way Random-Effects Model (ICC(2,  $K$ )) was selected, which models two effects: raters and ratees; and assumes raters ( $k$ ) were randomly selected from a population of possible raters and each contestant/performer was rated by  $k$  raters. Similarly, the contestants/performers are considered a random selection from a population of contestants/performers, allowing the generalization of findings to the larger population. The type of measurement chosen (mean of multiple raters –  $k$ ) assumes that the average of raters' ratings is considered in

rating each contestant's/performer's performance, and uses the "Average Measure" which computes an average of raters' measurements and compares the average to the different judges' average. The type of ICC chosen was absolute agreement, meaning that it measured how close in agreement the raters were in terms of scores. The results were then examined to determine the reliability of scores according to the guidelines for interpretation set forth by Koo and Li (2016).

Analysis of the ICC value also included analysis of the Confidence Interval (CI), which according to Koo and Li (2016), should be used to evaluate the basis of reliability. Based on Koo and Li's guidelines for interpretation, a 95% CI of the ICC value indicates reliability on the following scale: "Values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability" (2016, p. 161).

SPSS (23) was used to compute ICC along with their 95% confidence intervals. Separate ICCs were computed for groups of raters across nine competitions. This was based on mean ratings with  $k = 3$  and  $k = 4$  raters, absolute agreement, and two-way random-effects model.

Overall, the results of the ICC analysis showed high agreement on performance ratings among judges in all nine competitions. Though some research suggests that adjudication panel size may affect results (Bergee, 2003; Fiske, 1983; Fiske, 1975), the results of the present study were not affected by the number of judges, as seven contests utilized three judges and two contests used four. The ICC's for groupings of judges across all competitions ranged from .93 (West Jones Competition) to .99 (South Jones

competition), which indicates excellent inter-rater reliability within competitions using the Tyson Show Choir Rubric. Results also suggest that the rubric is a reliable measure that can be used by judges to evaluate show choir performance. Table 2 displays the results from each of the competitions, including the ICC value and CI for each competition.

Table 2  
*Competition and CI Results*

Intraclass Correlation Coefficient			
Competition	Intraclass Correlation <sup>a</sup>	95% Confidence Interval	
		Lower Bound	Upper Bound
Jackson Academy Show Choir	0.932	0.834	0.976
Invitational			
Jackson Prep Show Choir	0.938	.863	.977
Masters			
Mississippi Show choir Contest	0.978	0.950	0.992
Oak Grove Magnolia	0.976	0.941	0.993
Invitational			
Petal Show Choir Invitational	0.985	0.970	0.993
Purvis Show Choir Invitational	0.982	0.956	0.994
Robert C. Byrd Vocalfest	0.969	0.913	0.992
South Jones Invitational	0.986	0.970	0.995
West Jones Show Choir	0.927	0.850	0.970
Invitational			

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. Type C intraclass correlation coefficients using a consistency definition. The between-measure variance is excluded from the denominator variance.

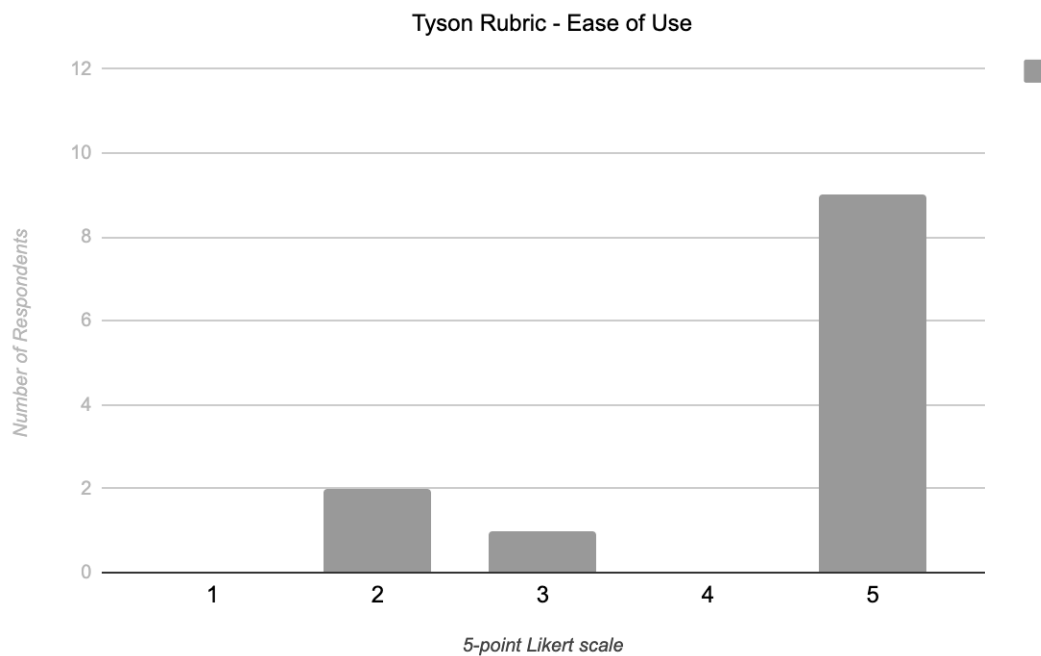
At the conclusion of the 2018-2019 competition season, all judges listed for each competition were sent a link to a survey via email. As noted in Table 1, the scores of 29 judges were statistically analyzed, which represents the preliminary competition results of each contest. However, each competition also hosted a Final Round of competition, which included additional judges and utilized an abbreviated version of the Tyson Rubric. Only preliminary scores were analyzed; however, all judges for each competition were asked to respond to the survey, even if each had only adjudicated the finals portion of a competition. The contest directors provided the email addresses for judges at each competition. A total of 32 judges were sent the survey link, while only 12 responded. All responses were anonymous.

The survey results determined that 25% of the respondents are currently employed music educators in a middle or high school choral program and 25% are music instructors at a college or university. The remaining 50% of adjudicators varied in occupation, including performers, choreographers, public relations, marketing, or human resource professionals. Teaching experience ranged from 0-5 years (25%), 5-10 years (33%), 10-15 years (8%), 15-20 years (8%), and 20 or more years (25%). Show choir adjudicator experience also ranged from 0-5 years (33%), 5-10 years (16%), 10-15 years (8%), 15-20 years (16%), and 20 or more years (25%), while 50% of adjudicators judge one or two contests each year, 33% judge five or more contests each year, and 17% judge three or four contests each year.

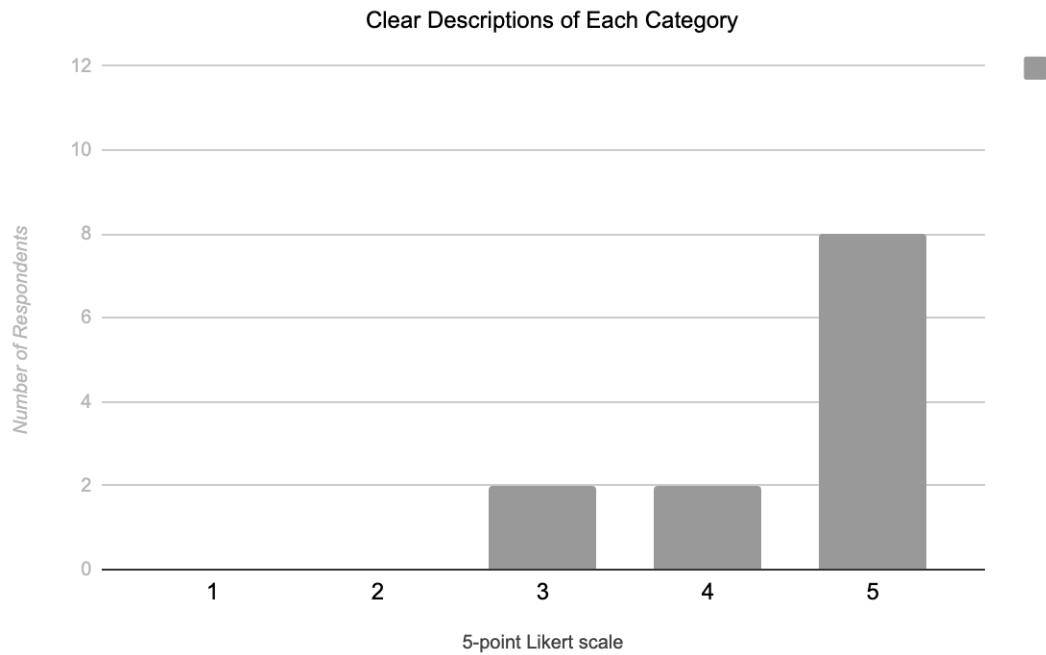
Prior to each competition, judges received an email from the researcher containing a brief orientation to the Tyson Rubric (see Appendix A), as well as a copy of the rubric. Half of adjudicators surveyed reported completing the Adjudicator Training

course, which was the email orientation. According to the survey results, 91% of adjudicators received information regarding contest rules, procedures, policies, or other orientation information prior to the contest in which each judged.

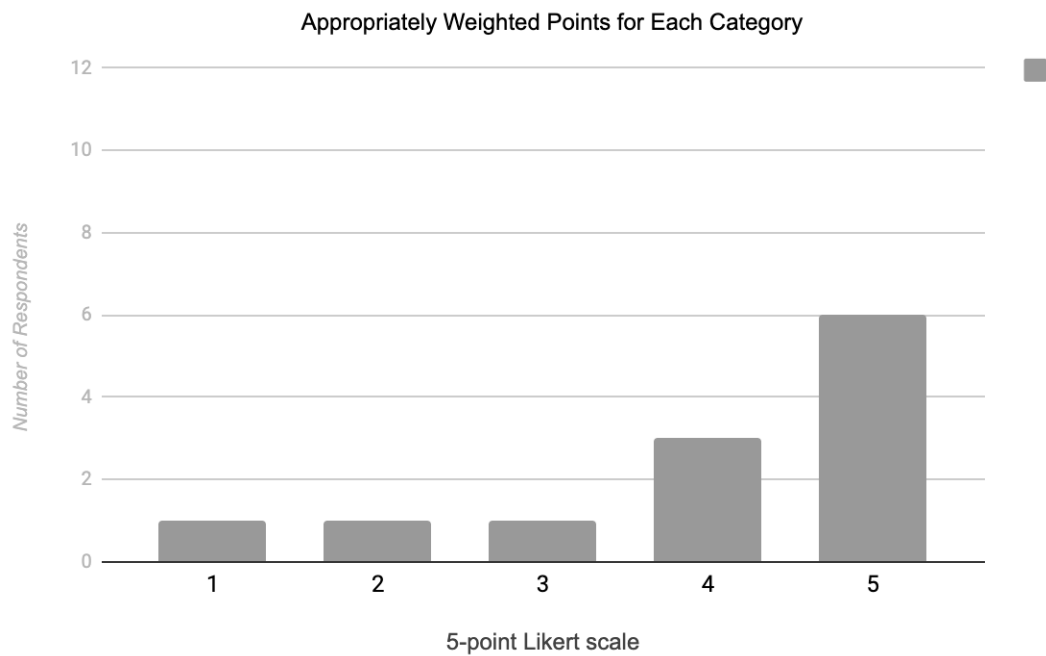
The remaining survey questions applied specifically to the Tyson Rubric as indicated in the Figures 4, 5, 6, 7, and 8. Adjudicators were asked to rate the Tyson Rubric according to ease of use, the clarity of descriptions for each category, appropriately weighted points for each category, how appropriately the rubric addresses the dimensions of a show choir performance, and the effectiveness for evaluating show choir performance. The responses were indicated using a Likert-type rating on a scale of 1-5, with 1 as poor and 5 as excellent.



*Figure 4. Tyson Rubric – Ease of Use*

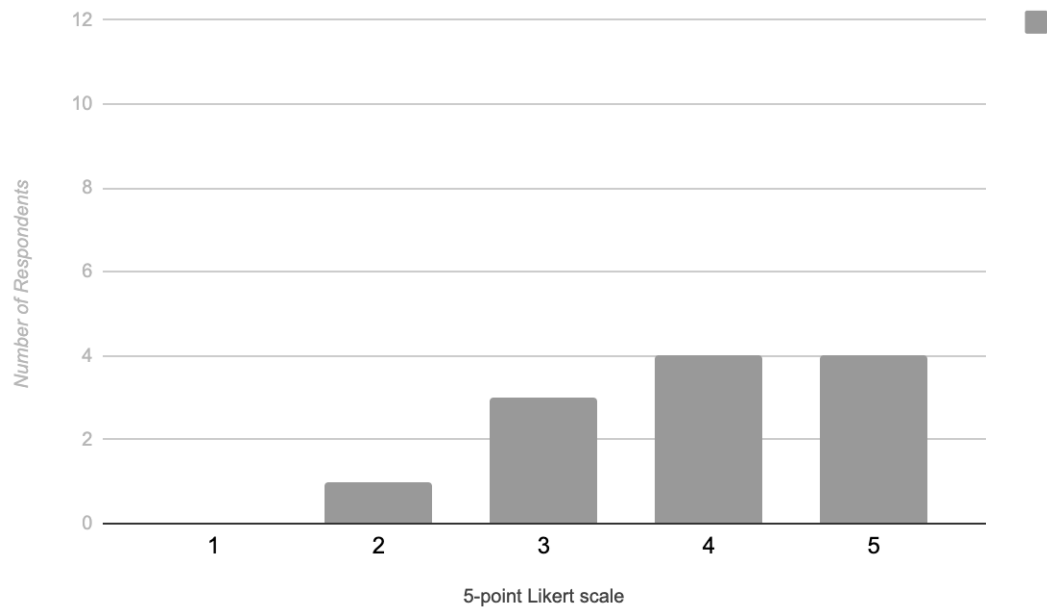


*Figure 5. Tyson Rubric – Clear Descriptions of Each Category*

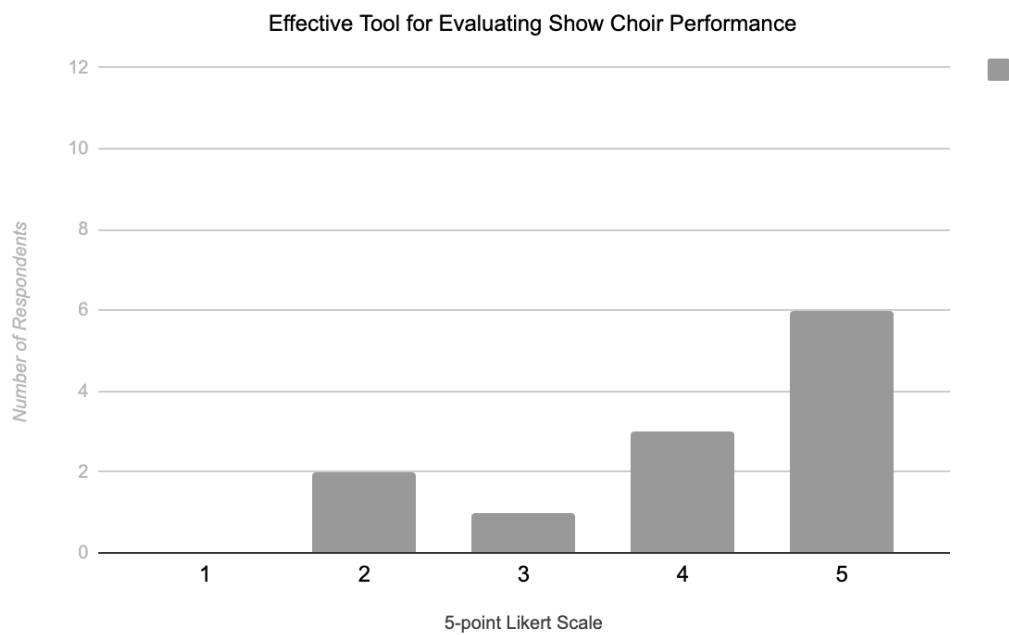


*Figure 6. Tyson Rubric – Appropriately Weighted Points for Each Category*

*Appropriately Addresses the Dimensions of Show Choir Performance*



*Figure 7. Appropriately Addresses the Dimensions of Show Choir Performance*



*Figure 8. Effective Tool for Evaluating Show Choir Performance*



The survey concluded with an open-ended question regarding suggestions for improvement to the Tyson Rubric, which yielded a variety of opinions by eight of the 12 survey respondents. Four judges indicated that a category addressing difficulty of choreography should be added to the rubric. In addition, one judge suggested that the visual categories should be more in-depth and reviewed to ascertain that the rubric includes all appropriate categories. Two judges suggested that a weighted point system is difficult to maneuver in application and suggested finding another system that gives weight to the overarching categories of vocal performance, visual performance, and show design, without using a weighted point system. Judges also suggested that it is difficult for one person to accurately judge all of the categories in the rubric, with possible solutions including fewer categories overall or caption judging. In this context, caption judging refers to adjudicators who only score the portion of the rubric that applies to that judge's area of expertise. For example, a choreographer may only judge the portion of the rubric dealing with visual performance and would not score a group's vocal performance.

### Qualitative Results

In addition to the quantitative survey, 10 judges participated in interviews with the researcher either via phone or video chat. Email requests for personal interviews were sent to the same 32 judges who received the survey link. All judges who responded affirmatively to the email request were interviewed. Prior to each interview, the researcher gained verbal consent from each participant to obtain, analyze, and publish the resulting data, while keeping the participants identities confidential. During the interviews, judges responded to 13 open-ended questions (see Appendix C), with

interview length ranging from approximately 13 minutes to approximately 35 minutes. Each interview was recorded, transcribed, and coded to determine emerging themes (see interview transcripts in Appendix D).

The adjudicators interviewed varied in occupation and show choir experience, with six identifying as current or former music educators either in high schools or at a higher education institution. One participant is a retired choral director with 30 years of teaching experience in a high school, and currently works as a consultant for a choir apparel company. Another participant currently teaches vocal music education at a public high school, while two others are currently vocal music instructors at higher education institutions. One participant is a former high school choral director who now works from home as a freelance adjudicator/clinician for show choirs, and another is a former band director with extensive adjudication experience with Bands of America and Drum Corp International who is currently employed in public relations and marketing at a higher education institution.

The four participants who do not identify as music educators have a diverse background of expertise. Three have experience primarily as choreographers and freelance adjudicators, one of which works full-time for an international cosmetics company. The final participant is a former show choir performer with experience in choreography, but has been employed full-time in Human Resources at a financial company for over 30 years. Of the ten interviewees, all but one had show choir adjudication experience prior to the 2018-2019 competition. All participants have been assigned a pseudonym to protect confidentiality.

The researcher emailed a brief explanation and copy of the Tyson Rubric to adjudicators prior to each competition. All but one of the adjudicators interviewed acknowledged reception of the communication, and affirmed taking time to become familiar with the rubric before arriving at the competition site. Additionally, all ten of the adjudicators described at least some form of on-site orientation that included an explanation of the rubric, allowed for questions regarding the rubric, and explained any other rules or procedures that applied to assessment.

*Bias in choral adjudication.*

Participants were asked to discuss the problem of partiality in choral adjudication, including the effect that a score sheet may or may not have on such bias. Studies indicate that nonmusical factors, such as performance time, school size, adjudicator experience and training, and order of appearance can influence performance assessment outcomes and cause bias effects (McPherson & Thompson, 1998; Napoles, 2009; Hash, 2013). McPherson and Thompson (1998) advise that bias may also occur if the adjudicator is familiar with a musician, in this case a performer or director, outside of the competition context. Referred to as the “halo effect,” the phenomenon is defined by Boyle and Radocy (1987) as “a positive or negative bias because of what the reader [judge] knows about the writer [ensemble]” (p. 123).

All participants interviewed acknowledged the existence of adjudicator bias in choral adjudication, though each interpreted the question in a different manner. Some spoke to a more general bias, with seven adjudicators indicating that the reputation of a particular ensemble or director often affects the outcome of an assessment. One participant, Mr. Walley, responded that

When you judge a show choir competition, you know who the best groups are before they even walk on the stage. And chorally speaking, if you've got a choir that has sung in recent years at a National ACDA Conference and you're judging their district MPA when you heard them at the ACDA Conference last year, you're going to sometimes judge on what you've heard in the past, not what you're hearing on that day. (personal communication, December 18, 2019)

Mr. Highland, a cosmetics professional with extensive experience in choreography and adjudication, made similar remarks, stating that

Because you have people who are familiar with a certain region who are working with certain groups in certain regions and so they're influenced by their experiences with those particular students, those particular choral directors . . . sometimes they are influenced by the name of the group." (personal communication, December 16, 2019)

A conflict of interest of this nature can be difficult to avoid given the financial constraints of hiring judges from different geographical regions beyond the contest venue. Ms. Alan, an experienced choreographer, states, "The unfortunate thing is, we are at the point that there are so many different competitions on every given Saturday that the number of judges to go around and put qualified judges can't be spread around enough" (personal communication, November 1, 2019).

Other participants interpreted bias in choral adjudication as more specific to personal preferences of adjudicators. An adjudicator's personal concepts of proper tone and appropriate style, either vocally or visually, is likely to influence performance assessment. One participant, a vocal music instructor, suggests that "there tends to be a bias on the part of the adjudicator because of whatever style they teach and prefer" (J. Bruser, personal communication, December 20, 2019).

The participants referred to several nonmusical factors that may lead to adjudicator bias, such as the scores and opinions of other adjudicators. At many contests,

judges sit in the same area, or even at the same table, making it easy to compare scores and discuss thoughts about particular ensembles. One participant refers to this as “table talk,” and suggests that the occurrence can influence the outcome of an assessment (A. Alan, personal communication, November 1, 2019). Mr. Byrd, an adjudicator with experience in marching band adjudication, further implies production value, impressive sets and costumes, and the size of a performing group affects an adjudicator’s assessment of the performance (M. Byrd, personal communication, October 22, 2019). Byrd’s view echoes findings by Burnsed and King (1985), which infer that larger groups often receive higher scores than smaller groups.

Several participants suggested that adjudicators feel comfortable with a personal scoring system that may oppose the categorical descriptions of a rubric. For example, an adjudicator’s opinion of what level of dynamic sensitivity is deserving of a seven might differ from what a seven reflects on a rubric. Adjudicators may struggle to abandon biased ideas of scoring while adhering to a particular rubric or system.

When asked if the type of score sheet used impacts bias, 70% of adjudicators indicated that appropriate score sheets can mitigate adjudicator bias. One adjudicator suggested that when using rubrics with specific criteria, judges are less likely to sustain bias, particularly partiality towards reputable groups, than when using global score sheets that simply ask for a single numerical rating (J. Walley, personal communication, December 18, 2019). Additionally, some participants proposed that the use of a specific rubric requires that judges give justification for the scores given, which again reduces the likelihood of bias.

However, two of the adjudicators did not believe that a score sheet itself has any impact on adjudicator partiality. For instance, Mr. Highland asserted the belief that no qualities inherent in any score sheet affects adjudicator bias, rather, “it really comes down to qualification of who’s sitting in that seat” (K. Highland, personal communication, December 16, 2019). Adjudicator qualification, as shown in previous studies (Fisk, 1975; McPherson & Thompson, 1988), may influence both the reliability of scoring procedures and the outcome of music performance assessment.

Information gathered from participants indicate that the Tyson Rubric, though statistically a reliable instrument for show choir, cannot ensure that bias does not interfere with the adjudication process. While the rubric presents a guideline for scoring that promotes justification of scores and consistency both within and between judges, contest directors must still strive to hire qualified and experienced judges, and provide training in correct procedures before each competition.

#### *Score sheets in show choir adjudication.*

Participants were asked to describe experiences with different types of score sheets used for judging show choir competitions. Except for one judge, who had no previous experience adjudicating show choir contests, all reported experience with global score sheets. In the context of show choir adjudication, global score sheets refer to those that include either numerical scales for each category that is judged, or a numerical total. For example, many score sheets include subcategories within vocal performance, visual performance, and show design. Global score sheets require that adjudicators circle a number, usually between one and 10, for each subcategory. Some global score sheets simply provide a numerical scale between zero and 100, on which judges circle the

number that best represents the overall performance without any sort of categories or descriptors that explain what that number represents.

In addition to global score sheets, five judges described using rubrics for adjudication. Of judges who have used rubrics, three cited the Indiana State School Music Association (ISSMA) rubric as an example of a commonly-known score sheet (see Appendix D).

Additionally, two judges described past experiences using the Bands of America (BOA) adjudication system which employs a type of rubric for scoring. Judges rate the performance on musical and visual categories on a score sheet that includes rating increments based on the frequency with which the performers achieve proficiency. The BOA adjudication system uses specialized adjudicators for caption judging, meaning judges only score a particular component within the marching band performance, rather than judging all aspects of the performance.

Participants described the Tyson Rubric as similar to some score sheets currently in use, though differing in specificity and thoroughness of category and rating descriptions. The chief difference between the Tyson Rubric and other score sheets, according to the participants, is the use of weighted point categories. Judges indicated that though most score sheets have more subcategories within the vocal performance area, signifying that the vocal total is worth more than the visual performance or show design totals, the Tyson Rubric is the only rubric currently in use that utilizes varying point ranges for each subcategory. Aside from this characteristic, two judges found it comparable to most of the score sheets currently in use. However, Mr. Southward, a

choreographer, performer, and adjudicator, found that the weighted system complicated the adjudication process (D. Southward, personal communication, October 15, 2019).

The researcher asked participants to discuss the appropriateness of the category weights, including what points or categories might be altered to make the rubric more balanced. Responses indicate a variety of opinions on this topic, with 50% of adjudicators advising that no changes are needed, 30% recommending revisions to the current categories, and 20% preferring that all categories be worth the same amount of points.

The judges who recommended altering the Tyson Rubric all stated that the categories of style, both visual and vocal, should be allotted more than the seven points currently designated. J. Walley also proposed that the dynamic sensitivity and staging/transitions categories, which also receive seven points, should be weighted more heavily (personal communication, December 18, 2019). Another adjudicator advocated changing all of the categories that receive seven points, saying that sevens are difficult to navigate in comparison to ten's and five's (A. Alan, personal communication, November 1, 2019).

Only two judges felt that the rubric should not be weighted and that all categories should receive the same amount of points. For M. Silver, the weighted points are “discretionary” and have caused some judges to “forget about them,” since weighted rubrics are not widely in use in states other than Mississippi (personal communication, December 18, 2019). Silver suggests possibly only having one or two of the most important vocal categories weighted more heavily, with the remaining categories all worth ten points. The other adjudicator advocating for equal weighting for categories suggested that if a category is not as important as others, perhaps the solution is to not



include it in the score sheet, rather than assign fewer points (D. Southward, personal communication, October 15, 2019).

### *Usability of the Tyson Rubric*

In terms of usability, participants varied in response, with half of adjudicators interviewed finding it cumbersome or daunting, and half stating that usability is reliant upon familiarity and training. Mr. Byrd and Ms. Bruser found that trying to score all of the elements of the performance was overwhelming within the time frame of a typical show choir performance (18-20 minutes), though both acknowledged that the problem was not necessarily due to the rubric (M. Byrd, personal communication, October 22, 2019; J. Bruser, personal communication, December 20, 2019).

A confounding factor in usability might be the use of Carmen Scoring Systems (CSS) in conjunction with the physical copy of the rubric. At each of the nine competitions, judges were given a physical copy of the entire rubric, which included the full descriptions and rating scheme. However, judges do not use the physical copy of the rubric for scoring, instead using laptops or iPads to score with CSS. Though CSS includes the correct titles and weights for each subcategory, full descriptors are not included, thus adjudicators must refer to the physical copy to ensure correct scoring procedures. While 30% of the participants commented specifically on the ease of use of CSS, one judge also mentioned the possible problems in having to score on one implement while referring to another. K. Highland, a choreographer with extensive adjudication experience, stated that it is “something that can subconsciously make you also disassociated with the numbers of the rubric if they’re not together” (personal communication, December 16, 2019). Future research is needed on this topic in order to

determine if performance assessment reliability is affected by scoring in a system that does not include a rubric in its entirety.

### *Benefits and limitations of the Tyson Rubric*

Participants identified numerous benefits associated with using the Tyson Rubric for show choir adjudication, predominantly its potential to improve educational feedback for performance assessment, as mentioned by 50% of adjudicators interviewed. Research indicates that the job performance of school music ensemble directors is often evaluated and determined, at least in part, by ratings of the ensemble at festivals and contests (Forbes, 1984; Rogers, 1983; Burnsed & Sochinski, 1983). For this reason, competitions must make every effort to ensure quality adjudication that provides meaningful feedback. Several participants propose that the Tyson Rubric allows for specificity of feedback that is useful for improvement in the choral classroom. Mr. Walley stated,

I think what we forget is the end game, the end result, needs to be information for the teacher and the students on stage that they can use to improve. And a lot of score sheets don't give that, but the Tyson rubric does because you've got those boxes. And I think a lot of adjudicators forget it's not just about a score, it's giving information so programs can improve. (J. Walley, personal communication, December 18, 2019)

M. Field suggested that using the rubric as a teaching tool for feedback and for preparation is similar to knowing the correct answers to the test, and further states, "I think it's really good for the classroom and for rehearsals for students about how to aspire to be a five out of five, instead of a four out of five" (personal communication, November 22, 2019). Similarly, Mr. Hayworth felt that the Tyson Rubric is useful for the growth of an ensemble and program development, because it allows for comparison between contests that use the rubric (personal communication, October 25, 2019). Thus, groups

can measure improvement throughout the three-month competition season. Another participant, J. Bruser, reiterated the educational benefits of the Tyson Rubric, particularly if more contests were to adopt the same rubric, as a way to implement standardization across the genre, and compares show choir contests to athletic competitions, saying,

If we think about it in terms of the fact that it is something that is done competitively, there's no other arena where you do competitions where you don't know what the rules are or you don't know what you're being measured on. Oh, you didn't stand on your head? Well, I didn't know that was part of it. So, with a good rubric and with a system in place that says these are the things and the qualifications you're going to be judged on, then you've leveled the playing field (personal communication, December 20, 2019).

Ms. Beech shares similar views linking the specificity and clarity of the Tyson Rubric to improved feedback, stating that the “clarity for all parties” allows adjudicators the freedom to communicate targeted feedback to a group without restricting points or awarding too many points (K. Beech, personal communication, October 16, 2019). Mr. Byrd also shared that the Tyson Rubric “clarifies scoring ranges, it clarifies what needs to be done” (M. Byrd, personal communication, October 22, 2019).

One participant, Mr. Highland, posits that the Tyson Rubric could be adopted nationally or regionally to promote better adjudication practices, stating “I like the specificity of what you're going for because I think it's educational, and I think it will encourage less bias” (K. Highland, personal communication, December 16, 2019).

In addition to more specific feedback, 60% of participants cited improved consistency with the use of the Tyson Rubric. Some of the judges refer to internal consistency, in this context meaning that the rubric allowed adjudicators to remain consistent and accountable in scoring throughout the day or contest without wide

fluctuations between ensembles. One judge, Mr. Field, a choral music educator, stated that the Tyson Rubric “helps me a great deal to keep me true and honest and away from bias” (M. Field, personal communication, November 22, 2019). Ms. Alan, a choreographer and freelance adjudicator, gave similar statements, stating that the Tyson Rubric keeps the judges more consistent with scores and prevents one judge from altering the outcome due to scores that vary wildly from the rest of the panel.

Several participants referred to the external consistency of the Tyson Rubric. External consistency, in this instance, is referring to the standardization of the rubric and the generalized consistency judges may gain when using the same sheet each week, as opposed to differing scoring implements each week. One participant stated “I just think the benefits of it are that it will lead towards consistency, and what’s being reviewed by the judges. I just think that’s a very critical thing, that . . . everyone’s on the same page about things” (M. Silver, personal communication, December 18, 2019). Mr. Southward echoed the idea, stating, “in my perfect world . . . everyone could use it so that there’s not a change week-to-week” (D. Southward, personal communication, October 15, 2019).

Participants were also asked to disclose any limitations of the Tyson Rubric, with 70% of judges stating that no limitations were detected. The remaining three participants described limitations in the physical format of the rubric and the use of weighted categories. In regards to the physical format, one judge in particular felt that the printed rubric was “tedious” because of the manner in which it was printed (J. Walley, personal communication, December 16, 2019). Contests utilizing the Tyson Rubric most often printed it on both sides of three sheets of standard copy paper, which meant that judges would have to flip pages to reference the rubric while scoring on CSS.

In comparable statements, Mr. Byrd, who has extensive experience adjudicating marching band contests, proposed changes to the formatting of the physical copy of the rubric by fitting everything on two sides of one sheet. The front side of the sheet might list the categories but only use “the meatiest words per caption,” while the back side of the sheet contains the full descriptors for each category (M. Byrd, personal communication, October 22, 2019). A further suggestion included caption judging to reduce the number of categories for which each judge is responsible.

Both Mr. Byrd and Mr. Walley allude to problems that may occur when 1) judges are not familiar enough with the sheet to score on a computer without constantly referring to the physical copy, and 2) judges risk missing vital elements of the performance while searching for the appropriate box on the physical sheet. This limitation is compounded, according to one participant, by the use of weighted categories. Mr. Southward, a freelance choreographer and adjudicator, stated that the weighted categories are limiting, particularly for judges accustomed to score sheets that award the same amount of points for each category in the rubric (personal communication, Oct. 15, 2019).

#### *Caption scoring.*

Many show choir contests utilize caption scoring, which emerged as a theme in discussions with the participants. Caption scoring refers to a panel of judges who do not rate all facets of the competition, rather each judge only adjudicates the categories that relate to personal expertise. For instance, a choral music educator with experience in vocal instruction only scores the categories under vocal performance, while a choreographer only scores the visual elements of a performance. The topic of caption scoring is widely debated among show choir professionals, with a wide range of opinions

on the matter. During the course of the interviews for the current research, six adjudicators discussed caption scoring.

The participants opposed to caption scoring included J. Walley, a retired choral director, M. Silver, a former participant and choreographer who works at a financial institution, and T. Hayworth, a vocal instructor at an institution of higher learning. Each expressed strong arguments against this particular type of scoring.

Mr. Walley's opinions stemmed from a career as a choral director whose groups competed at both regional and national show choir contests, saying, "Personally, when I've done that [caption scoring] I don't like it because it's like, okay, I'm only responsible for this, but in the classroom, I was responsible for all of it. To me, if you do that you're not using everybody's expertise" (personal communication, December 18, 2019).

Adjudicators may find it difficult to separate what is heard from what is seen, and even more challenging to only score or comment on one of those areas. Mr. Silver disapproves of caption scoring because of the restraints it places on adjudicators, saying, "When I've been asked to judge one of those two [vocals or visuals only] you're so, I hate to say myopic, but you kind of are, about what you're looking or listening to that you kind of lose sight of that overall thing" (personal communication, December 18, 2019). Mr. Silver goes on to describe experiences with a hybrid score sheet that allows vocal judges to score all of the vocal categories and a few of the most relevant visual categories, and vice versa for visual judges. Mr. Silver declared that the first preference for each adjudicator should be to judge the entire sheet or rubric, however "the really

critical asterisk [consideration] with that is finding the appropriately skilled judges to do that” (personal communication, December 18, 2019).

Financial constraints, as mentioned earlier, often prevent contest directors from hiring unbiased professionals who are qualified to judge all aspects of a score sheet. However, Mr. Hayworth posits that any judge qualified to adjudicate a show choir contest has the background and expertise to comment on any area of the score sheet, stating, “I have enough knowledge as a performer to know if something looks wrong. And just me judging on voice alone, you’re not utilizing your judge at the best of their abilities” (personal communication, October 25, 2019).

Utilizing caption scoring may be a more efficient way to adjudicate, regardless of a judge’s qualifications or area of expertise, as proposed by one participant. Mr. Byrd submits that adjudicating all areas of a show choir performance can be overwhelming and may dilute the quality of feedback, stating

I think the evaluation of the groups would be so much better, and at the end of the day, so much more helpful to the students and the teachers, if a movement professional were judging movement and nobody else was really in on that. If a design professional that does theatrical productions or shows of different things were doing a design show element, and that musicians focused on music . . . I think they’d get a lot better feedback. (M. Byrd, personal communication, October 22, 2019)

Byrd goes on to describe an accomplished and specialized adjudicator as “the extra teacher in the room,” who can provide feedback that leads to growth and improvement (M. Byrd, personal communication, October 22, 2019).

Utilizing a panel of specialized adjudicators, according to K. Highland, helps to not only mitigate bias, but provides a highly specific and “fine-tuned assessment of what’s being presented” (personal communication, December 16, 2019). Mr. Highland’s

point of view is that of a professionally-trained choreographer who can speak to the visual elements of performance, particularly style, in a manner that most chorally trained directors and educators cannot. Highland argues, “If you study a lot of style then you could speak to it. Now, would I expect a choral director who sings, plays piano and is really successful as a musician to understand some of the techniques in the style and dancing or choreography? Not so much. That’s not in their purview” (K. Highland, personal communication, December 16, 2019). Highland proposes utilizing caption scoring during preliminary rounds of competition, but switching to comprehensive judging for the final round of competition. Most show choir competitions include a preliminary round of competition for all ensembles, and then allow the five or six ensembles who score the highest points to continue on to the finals round held on the same evening. Mr. Highland proposes that employing a different adjudication strategy for the finals round may provide a more interesting and thorough competition (personal communication, December 16, 2019).

Mr. Southward shared similar thoughts regarding caption scoring, but believes the decision should rely upon the qualifications of the judging panel and indicates that if a contest hires “dual-trained” adjudicators, then a comprehensive scoring method fully makes use of each judge’s expertise. Mr. Southward, like Mr. Highland, also suggested that caption scoring and comprehensive scoring could be used as a way to make the finals round of competition more meaningful (personal communication, October 15, 2019). No consensus among judges was reached in the scope of this research, suggesting that the use of caption scoring for show choir competitions is a topic in need of further research.



*Closing remarks.*

Each interview concluded by asking participants to share any other thoughts or concerns relating to show choir adjudication. Though many discussed caption scoring at this point in the interview, other topics emerged as well, such as hiring and training qualified adjudicators. Though the topic recurred throughout each adjudicator's interview, two participants made specific mention of the issue in closing remarks. Mr. Walley stated, "And like I said before, this is a great judging sheet but it's only going to be as good as a person you have utilizing it. I think every competition host could do a better job of preparing the judge to be accountable and fair" (J. Walley, personal communication, December 18, 2019). According to Mr. Southward, hiring qualified adjudicators is a currently debated topic among show choir directors and adjudicators, asking, "should people be eligible to judge if they don't have formal training or qualifications for that, meaning, if someone doesn't have a degree or some kind of formal training and industry experience, should they be able to give feedback and comments on something. . .?" (D. Southward, personal communication, October 15, 2019). Some states, such as Indiana, require potential judges to complete a formal training course in order to be added to the list of approved adjudicators (Indiana State School Music Association, 2020). Further research is needed to determine current practices for adjudicator training in show choir competitions.

A former choral director, Ms. Beech, expressed a desire for unification of contest rules and procedures either regionally or nationally. A typical show choir competes at several contests within a three-month period, some within the same state or region and some in different areas of the country. Each contest employs different scoring methods

and operating rules and procedures, requiring show choirs to adjust on a weekly basis to prepare for the next competition. Ms. Beech states, “I think that since we’ve gone with unification in so many things, it might not be a bad idea to extend that to certain contest rules that we all agree should be in place” (K. Beech, personal communication, October 16, 2019). Specifically, Ms. Beech cites enforcement of time regulations as an example of disparity between competitions. Most show choir contests allow each ensemble 25 minutes to set up, perform, and exit the stage, and penalize groups for extending the performance over that time limit. For some groups that perform with elaborate sets and props, the time limit poses a significant threat for score reduction. Ms. Beech suggests that while most contests implement performance time limits, the rule is either not strictly enforced or not applied fairly to all ensembles (personal communication, October 16 2019).

The participant interviews reveal numerous issues to be addressed when examining show choir competitions and scoring, many of which have been discussed in research related to instrumental ensembles and choirs. However, research specific to the genre of show choir is limited and further studies are needed to determine best practices.

## CHAPTER V – Discussion

A mixed methods study was selected to ascertain if the Tyson Rubric, created to adjudicate show choir competitions, is a statistically reliable instrument for that purpose. The final chapter includes a discussion of the key findings as related to relevant literature on ensemble performance assessment. Also included are the implications for future research based on user perceptions of the Tyson Rubric.

A pilot study and personal experience was used to create a rubric for utilization in for show choir adjudication. Anecdotal evidence suggests that show choir competitions lack standardization in regard to scoring instruments and methods. The study sought to answer the research questions:

1. Is the Tyson Rubric a reliable instrument for scoring show choir ensembles?
2. What are the user perceptions of the Tyson Rubric regarding ease of use, categories and construct definitions, and the perceived educational value of the rubric as an instrument for feedback?

The Tyson Rubric was utilized in nine show choir competitions in a single competition season. The study utilized an intraclass correlation coefficient to determine inter-rater reliability of the Tyson Rubric. Koo and Li (2016) state that instruments must first be determined to be reliable prior to being used for assessment. Inter-rater reliability is “the degree to which different raters or judges make consistent estimates of the same phenomenon,” or the same performance behaviors (Frey, 2018, para.1). To determine inter-rater reliability, an intraclass correlation coefficient was used, which is a common approach accepted by statisticians (Koo & Li, 2016; Shrout & Fleiss, 1979). Because the contest judges were chosen from a larger population of possible adjudicators with similar

qualifications, the intraclass correlation coefficient was determined using the Two-Way Random-Effects Model, which is used when results are to be generalized to any raters who possess the same characteristics (Koo & Li, 2016).

The intraclass correlation coefficient for all contests ranged from .93 to .99, indicating excellent inter-rater reliability. The 95% confidence interval also suggests that the Tyson Rubric, when utilized as an assessment instrument for other show choir contests using different adjudicators, would yield similar results. As pertaining to research question one, the data seem to indicate that Tyson Rubric is a reliable instrument for scoring show choir competitions.

The second research question addressed user perceptions of the Tyson Rubric, which were determined by analyzing the results of a quantitative survey and qualitative interviews. The survey, sent to all adjudicators (n=32) for the nine competitions that utilized the Tyson Rubric, revealed conflicting views of the rubric from the user perspective, though only a small sample of adjudicators responded (n=12). Some of the information gathered in the survey also conflicts with responses given in the interviews. Conflicting views between the survey and the interview may possibly be attributed to response bias by interview participants who wished to give more positive views when personally contacted.

Survey results suggest that the Tyson Rubric is easy to use, though nearly half of survey respondents indicated otherwise when asked directly. Interviews yielded similar results, though some adjudicators indicated the rubric was cumbersome, daunting, or overwhelming. Ease of use, though not integral to the effectiveness of the rubric, may influence contest directors' choice of scoring mechanism (Brookhart, 2018). However,

the interview participants reported that ease of use increased with repeated experience with the rubric, as well as with the incorporation of some form of orientation to the rubric prior to initial use.

The descriptions of each category were determined to be clearly articulated by the survey respondents. Clarity of descriptions received the highest rating among survey questions, indicating that regardless of other disputed characteristics of the Tyson Rubric, the instrument itself is a clear and reliable assessment tool. Clear and explicit descriptors is a key element of successful rubrics (Brookhart, 2018; Stanley et al., 2002; Wesolowski et al., 2018).

The survey results indicate that the most contested attributes of the Tyson Rubric are the use of weighted point values and the categories included for assessment. In an examination of other score sheets used for show choir adjudication, many appear to equally weight each category, unlike the Tyson Rubric which assigns different point values to categories. Only half of the judges surveyed found the weighted points assigned to each category to be appropriate, and even fewer found the categories included in the rubric appropriate. In a similar study evaluating the use of a weighted rubric for performance assessment of large ensembles in Kansas, survey respondents disagreed both about the weighted values and the appropriateness of categories (Latimer et al., 2010).

The overall effectiveness of the Tyson Rubric for scoring show choir performances could not be decisively concluded from the survey results, though a majority of respondents indicated high levels of overall effectiveness. The disparity between adjudicators can be attributed to a variety of factors, including professional background, adjudicator experience, and personal bias. Choral music directors may have

different values in regards to performance aesthetics from adjudicators who freelance as choreographers or clinicians, an assertion which is corroborated by interview results.

The qualitative portion of the study included telephone or ZOOM interviews with ten adjudicators who used the rubric in the 2019 competition season. Interview requests were sent to the same list of adjudicators as the survey link. Since the survey respondents remained anonymous, it is not clear if those who gave interviews also responded to the survey, though some similarities emerged between the open-ended survey question and some interview responses. A slightly higher number of interviewees are current or former music educators than were the survey respondents, otherwise, the occupational backgrounds are similar in both cases. Throughout the course of the interviews, three main themes emerged: 1) the presence of bias in choral and show choir adjudication, 2) the need for standardized scoring, rules, and procedures at show choir competitions, and 3) the need for improved adjudicator training in show choir.

### Implications

The limited amount of research for the genre of show choir prevents meaningful correlations between the Tyson Rubric and other score sheets in use. However, high inter-rater reliability and positive user perceptions are consistent with similar studies in other genres, such as band, orchestra, and choir. Rubrics for large ensemble performance assessment have been found to be reliable instruments, improving both internal and external consistency among raters (Barry, 2009; Jonsson & Svingby, 2007; Latimer et al., 2010; Norris & Borst, 2007). Additionally, rubrics used for choral and instrumental ensemble assessment provide greater clarity and specificity in feedback for educational

purposes than Likert-type ratings systems (Latimer, 2013; Latimer, et al., 2010; Wesolowski, 2012). Results of the present study reveal similar findings.

High reliability and positive user perceptions, while a positive step towards more standardized and improved adjudication procedures, do not address many of the additional areas of concern in show choir scoring, such as bias in adjudication and the need for qualified and trained adjudicators. Stemler (2004) argues that regardless of the statistical reliability of an assessment rubric, inter-rater reliability is “the level of agreement between a particular set of judges on a particular instrument at a particular time” (p. 1). Thus, reliability is a situational concept that is reliant upon a host of other factors (Stellmack et al., 2009; Stemler, 2004). With this in mind, show choir contest directors should focus efforts on improving adjudicator training and hiring adjudicators with consideration to the relationships that exist between adjudicators and competitors.

Bias is a non-musical factor that affects choral adjudication, presenting itself in a variety of manners. The halo effect occurs when a group’s reputation, or previous performances, affects an adjudicator’s ability to score the group accurately at the time of performance (Fiske, 1977; Forbes, 1994; McPherson & Schubert, 2004; Schmalstieg, 1972; Schmidt, 2005). According to some of the adjudicators interviewed, the halo effect is one of the leading causes of bias in show choir competitions. Show choir is a specialized genre with a relatively small number of ensembles when nationally compared with choirs, bands, and orchestras, ensuring that the community of adjudicators, clinicians, and directors have a high level of familiarity and reputations are widely known. Thus, ensembles with a history of winning performances are often highly rated under the assumption of continued excellence.

Anecdotal evidence and information collected from the interviews even suggests that some show choirs have national name recognition, making it difficult to mitigate the halo effect. However, hiring adjudicators from outside of the state or geographic region in which the contest is held may help to moderate bias for most show choirs. Additionally, contest directors should avoid hiring adjudicators with personal or professional connections to a particular ensemble.

Interviews indicate that adjudicators also label personal aesthetic values as a form of bias encountered when scoring. Background experiences and training as related to areas such as choral tone, and vocal and visual style, influence how judges score those on a rubric (Forbes, 1994; McPherson & Thompson, 1998; Napoles, 2009). McPherson and Thompson (1998) state that characteristics inherent in each adjudicator, such as “personality, experience and musical ability, training in adjudication, familiarity with the performer, and familiarity with the repertoire” affect scoring (p. 15). Similarly, Napoles (2009) suggests that an adjudicator’s educational background impacts a person scores ensembles.

Hiring qualified adjudicators with appropriate experience may help negate the effects of personal bias on scoring. Additionally, 70% of the adjudicators interviewed suggested that the type of score sheet used may aid in removing personal bias. If categories include clear descriptions that require adjudicators to score with consistency and reliability, rather than preference, personal biases may be eliminated. Training is a valuable component relating to the effectiveness of scoring measures (East, 2009; Rezaei & Lovorn, 2010; Timmerman et al., 2011). A study by Brakel (2006) found that adjudicator training increased the reliability of the ISSMA scoring rubric, meaning that



training is valuable outside of user perceptions. Based on survey and interview results, most of the adjudicators in the present study received copies of the rubric via email prior to the contest they judged, with many describing additional orientation procedures at the contest site.

The second theme that emerged from the interviews is the need for increased standardization of scoring mechanisms, rules, and procedures throughout the competitive show choir community. Show choir is a nationally recognized performance medium, but is not subject to any sort of overarching governing body. The American Choral Directors Association (ACDA) does retain a repertoire and standards chair for Contemporary and Commercial Music, which includes show choir music (American Choral Directors Association, 2020). However, competition regulations and scoring procedures do not fall within the purview of the position, rather, the leadership role is limited to presenting relevant repertoire/resources at national conferences, encouraging standards for high quality performances, and coordinating the efforts of division chairs across six national regions. Similarly, state and division chapters of ACDA also include Contemporary and Commercial music positions, though no evidence exists showing that ACDA influences show choir competitions in any meaningful way.

The lack of oversight by a governing organization may be a problem in and of itself. Most of the show choir competitions are sponsored by high school choral programs as a fundraising effort, which, anecdotal evidence suggests, can be quite profitable. Additionally, some organizations, such as FAME, Heart of America, and Heritage Festivals, exist outside the realm of education and function as a for-profit competition series or travel agency (FAME Show Choir National Championship Series, n.d.; Heart of

America, n.d.; WorldStrides, 2020). Hash (2013) proposes that making festivals and competitions, upon which many directors and educators are evaluated, a place for profitable endeavors presents a conflict of interest. Hash states that

Permitting outside organizations to play a significant role in evaluation removes part of this process from the oversight of local school authorities—who ultimately are responsible for providing fair, reliable, and valid teacher appraisals—and creates a conflict of interest for contest sponsors by making them choose between providing accurate assessments and maintaining director participation. (Hash, 2013, p. 166)

Some states, such as Indiana and Iowa, have separate entities that govern show choir competitions, and thus have some uniformity in scoring regulations (Indiana State School Music Association, 2020; Iowa High School Music Association, n.d.). However, any standardization with state organizations is not generalized to competitions outside of that state. Many show choirs compete in different states, and even different regions, perhaps on a weekly basis, meaning that each week of competition differs in scoring procedures and competition rules.

Some competitions continue to use traditional paper score sheets, on which the adjudicators rate and write comments, while many have progressed to recording comments in real time and using the sheet only for marking scores. Still others, like those in Mississippi, utilize computer-based scoring. The current study examined contests in which evaluators scored on a computer-based system while referencing a physical copy of the adjudication rubric, which presented a challenge for some of the judges interviewed. Alternating between a screen and a sheet of paper may cause confusion and lead to inefficiency in scoring.

Interview participants indicated several suggestions for revisions to the Tyson Rubric. Both survey respondents and interview participants proposed the use of caption

scoring as a way to improve adjudication measures. Caption scoring refers to utilizing each judge only for a particular portion of the score sheet, rather than comprehensive scoring in which each judge scores all categories on the sheet. At the time of the current study, no research examining the effectiveness or reliability of caption judging versus comprehensive judging for show choir competitions was available. Yet, some studies indirectly address caption scoring for music performances, suggesting that final ratings are reliable regardless of the reliability of each category (Burnsed et al., 1985; Latimer et al., 2010).

Benefits of caption scoring could include increased specificity of feedback, as indicated by some of the interview participants. The assumption is that reducing the overall amount of scoring categories for which a judge is responsible would lead to more specialized and detailed feedback for the ensemble. Caption scoring could lead to improved ease of use, specifically with the Tyson Rubric. Survey and interview responses imply that the Tyson Rubric can be overwhelming or daunting due to the large number of categories on the sheet, which may be alleviated if judges are only responsible for a portion of the categories. Yet, because of the multiplicity of performance factors that may just be the nature of the show choir genre. The focus of education versus for-profit contests may completely determine caption versus comprehensive scoring.

Conversely, some interview participants rejected the notion of caption scoring, stating financial constraints and decreased quality of feedback. Employing caption scoring would require the contest director to find specialized judges in each area of the score sheet, including vocal performance, visual performance, show design/overall effect, and in some cases, accompaniment. Participants allege that it is difficult to find such

expertise in a small geographic region, meaning that contest directors would have to hire adjudicators from a broader, more distant region. Since contest directors are responsible for reimbursing travel costs associated with adjudication, hiring judges in this manner might prove financially restrictive. On the other hand, hiring from outside of the contest area, as discussed previously, would also help to eliminate scoring bias, negating the impact of financial constraints in favor of a more worthwhile, impartial practice.

Caption scoring, according to some of the interview participants, might also dilute the educational value of feedback. Rather than assessing a performance in a holistic manner, caption scoring reduces the scope of adjudication in such a way that judges may feel constricted in giving feedback. Additionally, one interview participant noted that the aspects of a show choir performance do not function in isolation, rather, it is a cohesive blend of all elements of performing and cannot be viewed independently. That is to say, the manner and extent to which a group executes choreography may have a direct effect on vocal execution, as well as the overall musical expression of text (Green, 1984; Liao, 2002). Likewise, the overall design of a show, which includes construction and pacing, is also affected by elements such as transitions, blocking, and accompaniment. To only score and comment on one aspect assumes that each is not intertwined.

In addition to caption scoring, users expressed difficulty in adjusting to the weighted point categories. While the initial pilot study indicated that a weighted rubric might place value on the most significant portions of a performance, users struggled with the weighted system. Contest directors may wish to use the rubric, but alter the point values so that each category is worth the same amount of points. It is unlikely that reliability will be affected if point values remained consistent throughout the rubric.

A weighted point system ultimately reflects the values of the creator of the rubric, even if the rubric is based on prior research or a collaboration of other contributors. The disparity of opinions regarding appropriately weighted point values of the Tyson Rubric may be attributed to the personal aesthetic values of those surveyed. Likewise, certain categories were requested to either be added to or deleted from the rubric, most likely based on the values of the respondent. Suggestions for more appropriate categories and weights skewed to the visual portion of the rubric. For instance, difficulty of choreography was the most suggested additional category. Respondents suggested exploring other procedures to weight a rubric, as opposed to using categories that all receive the same amount of points. Ridding the rubric of weighted points, according to the survey respondents, would make the Tyson Rubric easier to use.

In addition to scoring measures, rules and procedures also lack standardization across the genre, as noted by interview participants. Examples of differing rules and procedures include group classification and performance time limit, though research does not yet include literature on show choir rules and procedures. Evidence gathered in regard to operating rules/procedures is anecdotal and related to the personal experience of the researcher. Groups are classified as small, medium, and large, but that classification is dependent upon different factors at each contest. For example, some contests classify group size based on the maximum number of performers on stage at one time, while other contests use school population to determine classification, creating confusion and possibly unfair adjudication procedures. Classification based on school population may determine that a group is in the large division, even though there are only 20 performers on the stage, while a school of smaller demographics may have upwards of 60 performers

on the stage simultaneously. Performance categorization in the manner above greatly affects points awarded, as studies have shown that larger groups typically receive higher scores (Hash, 2013; King & Burnsed, 2009; Latimer, 2013; Napoles, 2009; Rogers, 1983).

Performance time limitations also have an effect on scores, and possibly outcomes, of show choir contests. At many show choir contests, ensembles are limited to 25 minutes of performance time, which includes set up, actual performing time, and striking the set/band. The rule is loosely enforced across competitions, though contests that follow the procedure typically enforce a penalty for exceeding the time limit. One judge suggested that it would be helpful for contests to adopt the same policy regarding time limitations and penalties.

In a genre that still faces stereotypes and concerns of legitimacy within the realm of choral music education, increased standardization and reliable scoring practices would lend credibility. Furthermore, standardization of practices would enhance the educational merits of show choir and promote positive learning experiences through competition.

The final theme that emerged from the interviews is the need for implementing and improving adjudicator training. Research states that adjudicator training is essential to both fairness in scoring and director perceptions of feedback (McPherson & Thompson, 1998; Rawlings, 2019; Stegman, 2009; Wesolowski, 2012). Yet, research conflicts when linking adjudicator training directly to increased reliability in scoring (Brakel, 2006; Fiske, 1977; Fiske, 1978). While some states incorporate adjudicator training for show choir contests, like other issues of standardization, no universally accepted course or method is in current practice.

The nine competitions that utilized the Tyson Rubric agreed to allow a brief orientation to the Tyson Rubric via email. Of the survey respondents and interview participants, 73% indicated receiving an instructional orientation email that included a copy of the Tyson Rubric, and nearly all reported additional training or orientation at the contest site. Without a control group of adjudicators who did not receive training, the study cannot conclude that training increased reliability or had a meaningful connection to positive user feedback. However, some of the interview participants indicated that receiving the rubric prior to the contest was important in navigating the rubric during the competition, facilitating better ease of use.

Improving adjudicator training across the genre of show choir would lead to greater standardization, though it might prove difficult in the absence of a national governing body. A more manageable solution might be to implement adjudicator training regionally, or among contests that share scoring procedures and rubrics. Like the ISSMA adjudicator certification course, orientation and training could occur rather easily via an online Learning Management System (LMS), like Canvas or Blackboard, though a financial commitment would be required for use (Indiana State School Music Association, 2020). Collaborating with computer scoring systems, like Carmen, to include rater training when used is another prescription for improved practices. The general lack of standardization in any area of show choir scoring and procedures implies that steps taken to initiate new measures would be an immediate improvement over current practice.

## Limitations

The study of the reliability and user perceptions of the weighted Tyson Rubric for show choir adjudication denotes many topics for further research. Yet, the extent to which the results can be generalized is unknown, due to several limiting factors, including the small sample size of qualitative results, the researcher's personal experience with the interview participants, and the localized sample of show choir contests.

The quantitative results of the study, which indicate high reliability, include a 95% confidence interval, meaning that it is highly likely that results would be similar if the study were replicated in similar situations. Therefore, no limitations in the statistical data were detected. Unfortunately, the survey and qualitative results differ greatly in generalizability, which is, in part, due to the individuality of subjects' responses to open-ended questions.

Though show choir is a nationally recognized genre of choral music performance, it does not pervade music education as compared to other large ensembles, such as choir, band, and orchestra. Financial constraints often prevent choral directors from including show choir in school programs. The addition of a costume or costumes, live band, backdrops/props, choreography, and the rising costs of arranging rights for show choir music contribute to the financial burden of choral programs already facing tight budgets (Burrack et al., 2014; Spohn, 2008; West, 2012). Although pockets of popularity exist in regions across the country, the community of show choir directors, choreographers, and adjudicators is relatively small, reducing the ability to achieve a large number of responses for qualitative research.



The relatively small number of qualified adjudicators and the financial constraints of hiring judges from outside of a particular region, due to associated travel costs, leads to repetitive regional judging from many adjudicators. For a show choir to be rated by the same adjudicator more than once per season is not an uncommon occurrence. Within the present study, adjudicators who were sent the survey link may have used the Tyson Rubric at more than one competition, but only provided one survey response.

Finally, the survey questionnaire and interview requests received a poor response from the total number of judges listed at the nine competitions using the Tyson Rubric. Though data from 32 judges was statistically analyzed, only 12 responded to the survey and 10 responded to requests for interviews, despite multiple attempts to contact the adjudicators. A clear explanation for the lack of participation is not readily available, though it is an example of sampling bias in qualitative research (Moser, 1951).

In addition to the small number of participants in the survey and interviews, personal relationships may be a limiting factor of the qualitative data. The study of the Tyson Rubric was born from a need perceived by the researcher's experiences as a show choir contest director and frequent adjudicator. As such, the researcher has worked with all of the interview participants to some extent, resulting in bias in the collected data (Moser, 1951). Adjudicators possibly experienced difficulty in stating opinions, particularly in questions pertaining to the Tyson Rubric, to the creator of the rubric. Though personal relationships are a limitation of the study, nevertheless, interviews produced significant topics for future research unrelated to the Tyson Rubric.

The final limitation of the study is the relatively small geographic region in which the study was conducted. Aside from one contest in West Virginia, all of the contests

used for research were in Mississippi. Show choir preferences, styles, and show designs differ greatly geographically. For example, California show choirs are known for extravagant productions centered around a main theme or retelling of a classic story, and must include an a cappella ballad. Midwestern show choirs have reputations for excellence in singing and traditional show designs, while Southern show choirs differ greatly from one school to the next in terms of style and production value. User perceptions of the Tyson Rubric may differ depending on the geographic region in which it is used.

Perhaps the most limiting factor of the study is the lack of other empirical studies examining show choir. Existing literature is almost completely relegated to that of opinion and recommendation, rather than quantitative study. Much like the previous discussion regarding a lack of standardization in the genre, empirical research is nearly absent. Though the Tyson Rubric was determined to be a statistically reliable instrument for show choir adjudication, other score sheets have not been examined for comparison. Therefore, the Tyson Rubric cannot be deemed more reliable than other score sheets in use. Given the dearth of empirical investigation of pedagogy and educational facets of show choir performance and evaluation, the research is at least a confirmation of the dire need for further research in the genre and an affirmation of the current study.

### Suggestions for Future Research

In order to fully understand the reliability and effectiveness of the Tyson Rubric, further research is needed. As previously stated, the lack of empirical research in the genre of show choir makes generalization difficult, but presents an opportunity for areas of research, such as the reliability of other score sheets, the effect of adjudicator training

on show choir scoring, the outcomes of caption scoring versus comprehensive scoring, and ways to mitigate bias in show choir adjudication.

For many states, show choir exists in isolation, not beholden to any overarching governance. Score sheets and rubric are, therefore, subject to frequent change and apart from analysis. Future studies might compare Likert-type score sheets with rubrics in order to ascertain which is the most reliable. Additionally, studies can examine differences between weighted and non-weighted rubrics, both in reliability and user perceptions. Once reliability is determined, studies may want to concentrate on perceptions of both adjudicators and show choir directors. Adjudicators examine the ease of use and clarity of score sheets, while directors speak to educational implications.

Numerous non-musical factors have been shown to affect ratings in large group performance assessment (Bergee & Platt, 2003; Norris & Borst, 2007; Wesolowski et al., 2015). Rubrics may help guide adjudicators to quality ratings, but cannot guarantee reduced subjectivity in regards to non-musical factors (Stellmack et al., 2009). Studies should examine how non-musical factors affect the outcomes of show choir competitions.

Show choir contests often occur over the course of an entire weekend, with performance from early in the morning to nearly midnight in some cases. Furthermore, some contests have an additional second round of competition for high scorers that occurs on the same day as the first round. Performance time, order of performance, and size classification are examples of areas that may impact contest results and are in need of further research.

Mitigating bias in show choir adjudication is an important topic for future studies. As previously discussed, name recognition and conflicts of interest often pervade show

choir judging panels. Typically, show choirs are introduced by an emcee before taking the stage to perform. Introductions include the name of the school, as well as directors, choreographers, and arrangers associated with each group. Under these circumstances, it is nearly impossible for adjudicators to ignore previously held assumptions about a competing ensemble. Studies could examine the reliability between two sets of adjudicators, one set that does not hear the introductory material and one set that listens to the entire introduction, and compare the outcomes between the two sets.

Conflict of interest is a more challenging problem to solve, given the aforementioned financial constraints associated with bringing in judging from other geographical regions. The contest director is charged with the responsibility of disseminating a list of all performing groups to the judging panel prior to the date of contest, and encouraging judges to recuse themselves if a conflict of interest exists. Additionally, show choir directors could be required to submit a list of all choreographers and arrangers when registering for the contest, so that contest organizers can be aware of possible conflicts when securing adjudication panels.

The effect of adjudicator training on show choir scoring is another area for future research. While numerous studies laud the positive effects of training for adjudicators, none of these are specific to the genre of show choir (Brakel, 2006; Stegman, 2009; Wesolowski, 2012). Beneficial information could be gained from conducting studies that examine the comparative reliability between judging panels with differing training experiences for a scoring rubric. Additionally, researchers might examine various methods for adjudicator training, including comparisons between on-site and virtual

training courses. Further research is also needed to determine the content used for such courses.

Finally, future research in show choir should include the topic of caption and comprehensive scoring. Results from the present study indicate that caption scoring is a topic of concern among show choir adjudicators. Studies that compare the reliability of results between caption scoring and comprehensive scoring in the same performance situation would add value to the discussion. A proposed study might utilize two judging panels on the same day of competition with the identical scoring rubric, but with one panel rating using the entire rubric and the other judging only a specified number of categories related to either vocal or visual performance. To date, no studies exist indicating differences or increased reliability between caption and comprehensive scoring.

### Conclusion

The study of the Tyson Rubric contributes to the narrow body of literature regarding practices in the genre of show choir by examining the reliability and user perceptions of a weighted rubric for adjudication. Statistical analysis, survey responses, and interviews with adjudicators indicate that the Tyson Rubric is a reliable and straightforward adjudication form with implications for continued use at show choir competitions. Furthermore, the promotes standardization of scoring practices. The study of the Tyson Rubric is an impetus for further research to continue developing sound procedures and scoring implements in the genre of show choir adjudication, and continued exploration of musical ensemble performance assessment practices.

## APPENDIX A – Tyson Rubric

Execution Of Choreography	<b>10 9</b> <i>Superior</i> Choreography is executed with synchronicity consistently throughout the show. Movements have a defined starting and stopping point.	<b>8 7</b> <i>Excellent</i> Choreography is executed well with few errors in sequencing and synchronicity.	<b>6 5</b> <i>Good</i> Choreography is decently executed with several errors in synchronicity. Sequencing is a problem in spots.	<b>4 3</b> <i>Fair</i> Choreography is executed well in some places but lacks consistent correct sequencing and synchronicity.	<b>2 1</b> <i>Needs Improvement</i> Choreography lacks synchronicity and sequencing is a prevailing problem.
Facial Communication	<b>10 9</b> <i>Superior</i> All members of the ensemble perform with facial expressions that are appropriate for the style and song selections/theme of the show consistently throughout the show.	<b>8 7</b> <i>Excellent</i> All members of the ensemble perform with appropriate facial expressions for style and song selections/theme of the show during most of the show.	<b>6 5</b> <i>Good</i> Some members of the ensemble perform with appropriate facial expression, but others do not do so in a consistent manner.	<b>4 3</b> <i>Fair</i> Appropriate facial expression is inconsistent among the ensemble and throughout the show.	<b>2 1</b> <i>Needs Improvement</i> Ensemble does not perform with appropriate facial expression.
Intensity	<b>7 6</b> <i>Superior</i> Ensemble dances with appropriate energy level and exhibits strength of movement throughout the show.	<b>5 4</b> <i>Excellent</i> Ensemble dances with appropriate energy level and strength of movement for most of the show.	<b>3 2</b> <i>Fair</i> The ensemble does not consistently demonstrate appropriate levels of energy and strength of movement (some members do and others do not).	<b>1</b> <i>Needs Improvement</i> The ensemble fails to demonstrate appropriate levels of energy and strength of movement.	
Staging/ Transitions	<b>7 6</b> <i>Excellent</i> Transitions and staging (timing between songs, blockings, entering and exiting the stage) throughout the show are well-planned and well-executed. Performers remain in character through transitions.	<b>5 4</b> <i>Good</i> Transitions and staging throughout the show are well-planned, with only minor issues in execution.	<b>3 2</b> <i>Fair</i> Transitions and staging are inconsistently executed throughout the show and are in need of further planning and rehearsal.	<b>1</b> <i>Needs Improvement</i> Significant attention is needed for transitions and staging.	
Style Appropriate to Music	<b>5 4</b> <i>Excellent</i> Dance style is appropriate for selections/theme.	<b>3 2</b> <i>Good</i> Dance style is inconsistently appropriate for selections/theme.	<b>1</b> <i>Needs Improvement</i> Dance style is not appropriate for selections/theme.		

Costuming	<div>5 4</div> <div>Excellent</div> <div>Costuming choices are stylistically and age appropriate. Costumes enhance the show design and are without distractions.</div>	<div>3 2</div> <div>Good</div> <div>Costuming choices are stylistically and age appropriate, but may be distracting at times (costume change issues, in the way of choreography, etc.)</div>	<div>1</div> <div>Needs Improvement</div> <div>Costumes are inappropriate for age and/or style. Costumes are distracting from the show.</div>		
Stage Conduct/ Professionalism	<div>5 4</div> <div>Excellent</div> <div>All members of the ensemble exhibit appropriate behavior entering and exiting the stage, and while the performance is continuing. Performers remain in character, acting and reacting to things that occur on stage (planned and unplanned) in an appropriate manner.</div>	<div>3 2</div> <div>Good</div> <div>Most members of the ensemble exhibit appropriate behavior entering and exiting the stage, and while the performance is continuing. Most performers remain in character, acting and reacting to things that occur on stage (planned and unplanned) in an appropriate manner.</div>	<div>1</div> <div>Needs Improvement</div> <div>Ensemble does not exhibit appropriate behavior entering and exiting the stage, and while the performance is continuing. Performers do not remain in character, acting and reacting to things that occur on stage (planned and unplanned) in an inappropriate manner.</div>		
Show Design					
Overall Effect	<div>10 9</div> <div>Superior</div> <div>Overall impression of the show is positive. Show displays superior attention to detail, creativity, and musicality.</div>	<div>8 7</div> <div>Excellent</div> <div>Overall impression of the show is mostly positive. Show displays much attention to detail, creativity, and musicality.</div>	<div>6 5</div> <div>Good</div> <div>Overall impression of the show is mixed. Show displays some attention to detail, creativity, and musicality.</div>	<div>4 3</div> <div>Fair</div> <div>Overall impression of the show is mixed. More attention is needed to detail, creativity, and musicality.</div>	<div>2 1</div> <div>Needs Improvement</div> <div>Overall impression of the show is negative. Lacks attention to detail, creativity, and musicality.</div>
Live Accompaniment Balance and Execution	<div>7 6</div> <div>Superior</div> <div>Live accompaniment is well executed and performed in balance with the ensemble throughout the show and contributes to the show's design and overall tuning.</div>	<div>5 4</div> <div>Excellent</div> <div>Live accompaniment is well executed and performed in balance with the ensemble for most of the show.</div>	<div>3 2</div> <div>Good</div> <div>Live accompaniment is inconsistently executed and often not in balance with the ensemble.</div>	<div>1</div> <div>Fair</div> <div>Live accompaniment is distracting from the show consistently in both execution and balance.</div>	
Recorded Accompaniment Balance and Execution	<div>Recorded accompaniment cannot receive a superior rating.</div>	<div>Recorded accompaniment is balanced with the ensemble and contributes as best as possible to the overall show design and tuning.</div>	<div>Recorded accompaniment is inconsistently balanced and does not contribute to overall show design and tuning.</div>	<div>Recorded accompaniment is distracting from the show in balance, show design, and tuning.</div>	
Concept	<div>5 4</div>	<div>3 2</div>	<div>1</div>		

Execution Of Choreography	<b>10 9</b> <i>Superior</i> Choreography is executed with synchronicity consistently throughout the show. Movements have a defined starting and stopping point.	<b>8 7</b> <i>Excellent</i> Choreography is executed well with few errors in sequencing and synchronicity.	<b>6 5</b> <i>Good</i> Choreography is decently executed with several errors in synchronicity. Sequencing is a problem in spots.	<b>4 3</b> <i>Fair</i> Choreography is executed well in some places but lacks consistent correct sequencing and synchronicity.	<b>2 1</b> <i>Needs Improvement</i> Choreography lacks synchronicity and sequencing is a prevailing problem.
Facial Communication	<b>10 9</b> <i>Superior</i> All members of the ensemble perform with facial expressions that are appropriate for the style and song selections/ theme of the show consistently throughout the show.	<b>8 7</b> <i>Excellent</i> All members of the ensemble perform with appropriate facial expressions for style and song selections/ theme of the show during most of the show.	<b>6 5</b> <i>Good</i> Some members of the ensemble perform with appropriate facial expression, but others do not do so in a consistent manner.	<b>4 3</b> <i>Fair</i> Appropriate facial expression is inconsistent among the ensemble and throughout the show.	<b>2 1</b> <i>Needs Improvement</i> Ensemble does not perform with appropriate facial expression.
Intensity	<b>7 6</b> <i>Superior</i> Ensemble dances with appropriate energy level and exhibits strength of movement throughout the show.	<b>5 4</b> <i>Excellent</i> Ensemble dances with appropriate energy level and strength of movement for most of the show.	<b>3 2</b> <i>Fair</i> The ensemble does not consistently demonstrate appropriate levels of energy and strength of movement (some members do and others do not).	<b>1</b> <i>Needs Improvement</i> The ensemble fails to demonstrate appropriate levels of energy and strength of movement.	
Staging/ Transitions	<b>7 6</b> <i>Excellent</i> Transitions and staging (timing between songs, blockings, entering and exiting the stage) throughout the show are well-planned and well-executed. Performers remain in character through transitions.	<b>5 4</b> <i>Good</i> Transitions and staging throughout the show are well-planned, with only minor issues in execution.	<b>3 2</b> <i>Fair</i> Transitions and staging are inconsistently executed throughout the show and are in need of further planning and rehearsal.	<b>1</b> <i>Needs Improvement</i> Significant attention is needed for transitions and staging.	
Style Appropriate to Music	<b>5 4</b> <i>Excellent</i> Dance style is appropriate for selections/ theme.	<b>3 2</b> <i>Good</i> Dance style is inconsistently appropriate for selections/ theme.	<b>1</b> <i>Needs Improvement</i> Dance style is not appropriate for selections/ theme.		

Costuming	<b>5 4</b> <i>Excellent</i> Costuming choices are stylistically and age appropriate. Costumes enhance the show design and are without distractions.	<b>3 2</b> <i>Good</i> Costuming choices are stylistically and age appropriate, but may be distracting at times (costume change issues, in the way of choreography, etc.)	<b>1</b> <i>Needs Improvement</i> Costumes are inappropriate for age and/or style. Costumes are distracting from the show.		
Stage Conduct/ Professionalism	<b>5 4</b> <i>Excellent</i> All members of the ensemble exhibit appropriate behavior entering and exiting the stage, and while the performance is continuing. Performers remain in character, acting and reacting to things that occur on stage (planned and unplanned) in an appropriate manner.	<b>3 2</b> <i>Good</i> Most members of the ensemble exhibit appropriate behavior entering and exiting the stage, and while the performance is continuing. Most performers remain in character, acting and reacting to things that occur on stage (planned and unplanned) in an appropriate manner.	<b>1</b> <i>Needs Improvement</i> Ensemble does not exhibit appropriate behavior entering and exiting the stage, and while the performance is continuing. Performers do not remain in character, acting and reacting to things that occur on stage (planned and unplanned) in an inappropriate manner.		
Show Design					
Overall Effect	<b>10 9</b> <i>Superior</i> Overall impression of the show is positive. Show displays superior attention to detail, creativity, and musicality.	<b>8 7</b> <i>Excellent</i> Overall impression of the show is mostly positive. Show displays much attention to detail, creativity, and musicality.	<b>6 5</b> <i>Good</i> Overall impression of the show is mixed. Show displays some attention to detail, creativity, and musicality.	<b>4 3</b> <i>Fair</i> Overall impression of the show is mixed. More attention is needed to detail, creativity, and musicality.	<b>2 1</b> <i>Needs Improvement</i> Overall impression of the show is negative. Lacks attention to detail, creativity, and musicality.
Live Accompaniment Balance and Execution	<b>7 6</b> <i>Superior</i> Live accompaniment is well executed and performed in balance with the ensemble throughout the show and contributes to the show's design and overall tuning.	<b>5 4</b> <i>Excellent</i> Live accompaniment is well executed and performed in balance with the ensemble for most of the show.	<b>3 2</b> <i>Good</i> Live accompaniment is inconsistently executed and often not in balance with the ensemble.	<b>1</b> <i>Fair</i> Live accompaniment is distracting from the show consistently in both execution and balance.	
Recorded Accompaniment Balance and Execution	<b>Recorded accompaniment cannot receive a superior rating.</b>	Recorded accompaniment is balanced with the ensemble and contributes as best as possible to the overall show design and tuning.	Recorded accompaniment is inconsistently balanced and does not contribute to overall show design and tuning.	Recorded accompaniment is distracting from the show in balance, show design, and tuning.	
Concept	<b>5 4</b>	<b>3 2</b>	<b>1</b>		

(Creativity of design)	<p><i>Excellent</i></p> <p>Show design exhibits great creativity and is appropriate for age level and group size.</p>	<p><i>Good</i></p> <p>Show design shows some creativity and is mostly appropriate for age level and group size.</p>	<p><i>Needs Improvement</i></p> <p>Show design lacks creativity.</p>		
Construction (pace of show)	<p><b>5 4</b></p> <p><i>Excellent</i></p> <p>Show design exhibits planning and great attention to pacing and energy.</p>	<p><b>3 2</b></p> <p><i>Good</i></p> <p>Show exhibits planning, but some attention to pacing and energy is needed.</p>	<p><b>1</b></p> <p><i>Needs Improvement</i></p> <p>Much attention to pacing and energy is needed.</p>		



## APPENDIX B – Adjudicator Interview Script

Hello! Thank you for taking the time to answer a few questions regarding your perceptions of the Tyson Rubric. This interview is being recorded, and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential. By continuing on, you are giving your consent to the interview and publishing of the resulting data.

1. What is your current occupation?
2. Could you briefly describe the types of score sheets you have used at show choir contests?
  - a. Are they rubrics, global scores, weighted, etc?
  - b. what do you think about the total points available for scoring? Is it too wide, too shallow to provide adequate/fair judging?
  - c. do evaluation sections need to be weighted? (should vocals and visuals be worth the same amount of points)
3. Could you speak to the problem of bias in choral adjudication?
4. Does the type of score sheet used have any impact on adjudicator bias?
5. Could you speak to the usability of the Tyson Rubric, from an adjudicator's point of view?
6. How does the Tyson Rubric compare to other score sheets you have used?
7. What orientation or training did you receive for the Tyson Rubric prior to the contest you judged? This could include receiving a copy of the rubric prior to the contest.
8. Are there any benefits of the Tyson Rubric for show choir adjudication?
9. Did you detect any limitations of the Tyson Rubric?
10. Did you think that the categories of the Tyson Rubric were appropriately weighted?
  1. If not, what weights might be more effective in giving appropriate balance of importance?
11. Are there any other thoughts concerning the use of the Tyson Rubric?
12. Are there other issues concerning show choir scoring that should be addressed in a rubric?
13. Tell me any other thoughts you might have about the Tyson Rubric, or show choir scoring in general, either positive or negative.

## APPENDIX C –Interview Transcripts

### Appendix C.1: “Mr. Walley”

(I = Interviewer; W = “Mr. Walley”)

I: Thank you for taking the time to answer a few questions regarding your perceptions of the Tyson rubric. This interview is being recorded and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential. By continuing on, you are giving your consent to the interview and publishing of the resulting data. Does that sound okay?

W: Absolutely.

I: Okay, my first question is, tell me your current occupation and then if you could follow that up with your background and expertise, experience in the show choir world.

W: Sure, I am now consultant, working as consultant, for Southeastern Performance Apparel. I do workshops and convention work for the company. I taught for 30 years in Georgia and Alabama, a combined 30 years. I was... Let me see... President of the Alabama Vocal Association and the Alabama Music Educators Association. We had, were trying to have, a well-rounded choral program with women's choir, concert choir, men's choir mentor and show choir. I've attended Showchoir Camp for 30 years worked as a clinician there for 15 years, competed probably 27... 28 of those years and did a lot of adjudicating across the country.

I: Okay, great, could you briefly describe the types of score sheets that you have used at show choir contests?

W: Yeah, when I first started judging it was basically the MENC judging guidelines, which was basically you had a number from 100 at the top, down to zero at the bottom and it was divided. 100-90 was a x score, and 89-80... was or whatever they did, but there was no... it was just a number, you randomly chose a number. That's where my adjudicating started with a form like that, and it really wasn't broken down choreographically, visually, instrumentally, vocally, anything like that. We just basically a vocal score sheet. And then it went on to a mainly vocal sheet, where you had to put a vocal score down and then it would be 10 points for choreography. It wasn't really a thought-out score. I've also used the... Oh, was it the BOA, Bands of America? That's the first time, I think was the Showstoppers choral competitions. You've heard of those, or...

I: I have, yes.

W: yeah, that's the first time I used the box scoring. Where you... and that made sense to me, it was a lot more difficult than yours. It was a lot clunkier. But the way that Showstoppers did it, you had a mirror judge on the opposite side of the theater, then you matched up. There was a vocal judge on the left, like vocal, choreography, instrumental, and on the other side of the theater, vocal and your scores had to match... not perfectly, but they had to be in the same general box as your mirror judge on the other side. So, that's kind of weird.

I: Yeah.

W: Yeah, and they would come back you and say, "Okay this judge saw it this way. We need to kind of come to a consensus". So they would have us change scores all the time,

to kind of be on the same, same wavelength as your mirror judge. That was kind of... that was kind of weird. And I think that was the first time I judged Showstoppers in Chicago in, I don't know the early '90s, and they only mentioned that once to me, so I don't know if people complained, but Ron Hellums who is that started that competition wanted the mirror judges to be in the same ball park scoring-wise.

I: That's bizarre.

W: Yeah, it really was, but I think they kind of dropped that notion after the first couple of years that I judge anyway, and I never heard it again but my first time there, that said, if you're going to be our judge you're kind of going to have to be in the same ball park. And I judged Darryl Ussary's competition when he was at Homewood High School. And he is a former band guy and he kind of used a band score sheet, but just changed it up a little bit and added some choral elements, but it was a big visual sheet and an instrumental sheet and a vocal sheet. It was kind of clunky, because there were a lot of points for each one. It wasn't just 100-point score sheet. And I've used several instruments like that. You'll have 250 points to distribute, which was kind of difficult.

I: Because it was too many points, you think like it was just too many points to keep up with?

W: Yeah, you could have a say you had 250 points total and your winner could have 230 points and your second place could have 150. There is just such a wide range of numbers between... And again, it was, any kind of sheet like that to me, where it was just random, where they asked you to just pick a number, for the vocal score and there's no rhyme or reason for why you chose a 93. You know, so.

I: Okay, wow, that's a lot of great information. Could you, and there are a couple of different ways that you could interpret this next question, so I'm not going to give you a lot of guidance, I just kind of want to hear your first thoughts. Could you speak to the problem of bias in choral and show choir adjudication?

W: Yeah, I think that's something that's been with us and will continue to be with us because when you judge a show choir competition, you know who the best groups are before they even walk on the stage.

I: because of reputation?

W: Yeah, because of reputation. And I'm not sure if that's where you wanted me to go. Say Petal High School or Clinton High School, or somebody like that, that walks on the stage, you kind of, and I've been guilty of this, you said, "Oh okay, well then well they're going to win anyway, so let me make sure my numbers match. You've got that bias and you like, okay let me match my numbers to what my perception is of that group not today, but in years past. And that group may walk on the stage and have a bad day. But they're still going to get, they're going to get scored because of who they are, I think. A lot of times.

I: I agree.

W: And chorally speaking, if you've got a choir that has sung in recent years at a National ACDA Conference and you're judging their district MPA when you heard them at the ACDA Conference last year, you're going to sometimes judge on what you've heard in the past, not what you're hearing on that day. And I think that's something that needs to be approached like, in every competition. These are the groups that are here, we're judging on what you see today, and you hardly hear anybody say that in an adjudicator's

meeting because I think everybody is biased for certain groups or what they've seen and heard in the past, instead of saying, "Okay today's today, and let's go just on what we see on stage today.

I: Do you along those lines, do you... does the type of score sheet used have any impact on that bias?

W: Absolutely, yeah. For example, if you just use a score sheet like the old MENC 0-100, that's going to favor your bias towards those groups. If you have a more in-depth score sheet where you can say, "Okay in Box 5 or whatever this box is, they didn't meet the criteria for this box. So I'm going to move them down to this box. You have certain things to look at where you can actually take points away in a reasonable manner, that if they didn't meet the criteria for this box they go to this box. Yeah, I think the score sheet is key.

I: Okay, so for the purposes of this interview and my dissertation the rubric that was used in Mississippi last year is being referred to as The Tyson rubric. It will not always be referred to as that and I have raised complaints about it but I'm required to call it The Tyson rubric. So, just so you don't think I have a big head. My committee is telling me I have to call it this for now. Could you, from an adjudicator's point of view, could you speak to the usability of the Tyson rubric, the rubric that is used in Mississippi?

W: The first time I used it, it was, it was daunting a little bit. It's a lot to look at, but after you've used it a group or two, you kind of fall in. And I think you have said this before we started the competition, when you... If you judge the boxes. If you're looking at the group and you put them in the correct box, use the sheet and it's going to help you and it does. It makes the the adjudicator be accountable because you want to be fair to the director and the students on the stage and you don't want to just pick a random number out of thin air. You've got those boxes to look at and say, okay, well, they're a box, whatever this box is, they meet the criteria. So let's put that there, and it makes you actually be accountable for the job that you're doing both for the competition host and the schools and directors that are being adjudicated. It makes you be accountable.

I: Great, and you've kind of touched on this a little bit, so you may not have a lot to say, but how does the Tyson rubric compare to the other score sheets that you have used? And it seems like if there's been a score sheet out there then John Baker has used that score sheet at some point, but how does this compare to those?

W: I don't know if this is the right way to put it, but it's more intellectual. It makes you think, as opposed to... I had somebody tell me one time, and it may have been Ron Hellums, he said, "You're a vocal judge, you're judging he said, you can actually put your head on your desk, close your eyes and just listen, and then put a score in a box." Which is... I'm like, "Well no, because it's all... It has to be cohesive everything has to fit together". And if they're doing choreography that doesn't allow them to be successful vocally, that's all part of it, right? What is the question again?

I: How the Tyson rubric compares to the other score sheets.

W: Oh yeah, and it's a... It's specific, as opposed to random.

I: What orientation or training did you receive for the Tyson rubric prior to the contest that you judged last year? This could include receiving a copy of the rubric prior to the contest.

W: Yeah, that did happen. I'm sitting here looking at the email right now where you sent me a rubric ahead of time, a couple of times a couple of years we've done that. And before we started, you gave us specifics on don't pick random numbers, look at the boxes and make sure you place the groups in the correct box for what they've achieved that day. Not in years past but that day at that time slot.

I: Okay, are there any benefits of the Tyson Rubric for show choir adjudication?

W: You've got to stay awake, you've got to be mentally alert, the entire time as opposed to some score sheet at the end, you just got to think one number. But with the Tyson rubric, you have to be alert, which is what we're paid for to be there, we want to be fair and we want to give a proper, a proper assessment of what's happening on the stage, and to do that, you've got to follow the certain criteria and put groups in that where they belong on that certain day on that certain time.

I: Okay, did you detect any limitations of the Tyson rubric?

W: The only problem that I had was, and I don't know if there's a way to fix this or not, LaDona, but when we have the sheets you got... I wish there was a big legal size sheet that had everything on one page as opposed to having to flip pages. It got kind of tedious a little bit because you wanted to do a good job, but you were like flipping pages, okay, where am I? Okay, yeah, I don't know if that makes sense.

I: Yeah, we've had... That's kind of been a recurring theme. And also, if there was a way that it could be when you're scoring in Carmen, if you could see it in when you see that, and I'm not sure that that's possible, but yeah, that's one of the big feedbacks that we've heard was to make it a little more, or a little less cumbersome, just in how we use it.

W: The first time I used it, I kind of wanted, and I don't know if you get this from other adjudicators, I kind of wanted to fight the score sheet a little bit.

I: In what way?

W: I wanted to do my own thing to how choose to do it. I said, Okay, they're kind of in-between boxes or I'm going to give them this number because this is what I think even though it's not on the sheet. You kind of want to fight it. But once you've done it a group or two, you kind of fall in line and you say, okay this does work. I don't know if you get judges that want to just fight the system and do it their own way. And I guess because you do have to work, you do have to be alert and you do have to be accountable, and a lot of competitions, you know how they are in Mississippi and across the country, they're going to start at 6 o'clock in the morning, and go to midnight, and that last group is like I'm tired of the boxes. I want to fight the system and do it my own way just so I can get finished quicker.

I: And there are things... And a portion of what I'm going to talk about that's come up that aren't in my results and have nothing to do with my rubric but through interviews and just looking at the different surveys and stuff I've done. There are a lot of things that no score sheet can fix. And part of that is qualified adjudicators. And then also like schedules like you said, a time of day that you perform affects adjudication. If judges are hungry, if they've been going all day, the order in which you perform because you kind of expect that the better groups go later in the day so you reserve points. So there are a lot of things that no score sheet can fix. So that's some things that I'll be talking about in the discussion portion.

W: Well, I think what we forget about as adjudicators, we're concerned about that day and being accountable to the score sheet, but I think what we forget is the end game, the end result, needs to be information for the teacher and the students on stage that they can use to improve. And a lot of score sheets don't give that, but the Tyson rubric does because you've got those boxes. This is what I didn't meet, I want to meet this box, that's the end game. I want to move from this box to this box. And I think a lot of adjudicators forget it's not just about a score, it's giving information so programs can improve.

I: Okay, so the next question is a little different. Did you think that the categories of the Tyson rubric were appropriately weighted, and if not, what weights might be more effective in giving appropriate balance of importance as far as the different categories?

W: Let me say I'm trying to... I had that pulled up on to it. What are the basic areas...

I: So, of course, it's more heavily weighted vocally. There are more vocal points than there are visual or show design, but then even within say, vocal performance categories like intonation and tone quality get 10 points and then dynamics, for example, is worth 7, and then difficulty of music is worth five. So within each broader category, the sub-categories are weighted as well.

W: Yeah, I pulled that up, I've got it right here. Starting with vocally, I like that you went with 10 points tone quality and intonation and diction. To me, dynamic sensitivity could actually be given up to 10 points. I think it goes from one to 7.

I: Yeah, it does.

W: Is what I'm thinking. Because you got excellent and then good. I don't know if you could do superior, excellent, good, needs improvement like the other vocal. It was just... To me, it would lend a little more weight. And you know, the style/interpretation I think might could use more points.

I: Is that one worth five?

W: It's a five. Because you think about it, what did Mark Myers say? There is only one way to sing correctly, but there are a gazillion styles of what you can sing. And maybe that's not enough points for style and interpretation because that's a huge part of the show. Yeah, there was another one, staging and transitions in the visual.

I: That's worth less points, yeah.

W: That goes one to seven because that's one that is really big visual. The actual choreography is big, but if they don't get the staging and transitions right, the choreography, the movement, which is worth the execution ten points, you're not going to be able to see the execution if the staging/transition isn't right.

I: One of the other, and I'm interested in your opinion, one of the things I've heard a lot is, especially from visual judges is that they wish there were a difficulty of choreography category and I've heard both sides of that and I was wondering how you feel about adding that.

W: Yeah, you've got intensity, which needs to be its own thing. Yeah, but how many points would you give it and how do we know how difficult it is? May it may look easy to me, or may look difficult to me, but it's easy for somebody else, you see what I'm saying?

I: I do. And then there are groups that aren't capable of more difficult, like it's, it's really subjective, what's difficult and what's not and how that relates to each group. And maybe you've got a group that's doing choreography that's difficult for them and it's very clean

but there's not a lot of dancy moves, and then you've got a group that just happens to have a lot of great dancers. And it's difficult to quantify, I think.

W: Right and you have to be extra accountable because you've got to look, you gotta look at the top level, and the level on the floor because the great dancers are going to be on the floor and they're going to be doing it great, even though it's difficult. But the people on the back right corner, that's where you're going to have to look to see. I kind of look at that intensity. It's not the same thing as difficulty, but maybe they're kind of related, a little bit.

I: Yeah, yeah, it's just been interesting to hear that I just wanted your opinion on that. That one's been popping up from the choreography specific folks.

W: Yeah, and did that come from choreographers?

I: Yeah, no, it came from people who were visual in their expertise. Centered on choreography.

W: Yeah, but like you said, if you can have simple choreography, like Marty DeMott choreography, and it's clean as a whistle I'd rather have that than something that's difficult, that half the kids can't do, but the front row looks fantastic.

I: Okay, do you have any other thoughts about the use of the Tyson rubric specifically?

W: I like how you address live accompaniment and recorded accompaniment, because I've been to so many places where it's like, okay, they use a tape... must give them a three. What if it's a fantastic recording? And they took time to do it or actually went into a studio, had people recorded it because they couldn't use live musicians. You give credit for that if it's done well. Because a lot of times you just kill a group just because of using a recorded accompaniment as opposed to live accompaniment. So I like how that is a positive and I think you want to weigh on the side of the students. Well, it's a middle school because they couldn't use instruments, but you don't want to destroy their opportunity to be successful.

I: Are there other issues concerning show choir adjudication, show choir scoring in general, that should be addressed in a rubric or scoresheet that are not currently addressed.

W: Oh, I'm not sure what word I would use for this. Appropriateness of movement... Appropriateness of music selection. A lot of times you just end up writing a note to the director, which I guess, is where it would be, but there may be an elementary school group that sings "Hey, Big Spender", stuff like that. You have style appropriate to music, but it's musical selections appropriate to the age.

I: Oh, that's great, nobody has said that yet, John, I totally agree that's great. The last question that I have is just any thoughts that you have in general about show choir adjudication, rubrics, scoring, judges, anything else that we didn't talk about or discuss that you think might add to the discussion?

W: And like I said before, this is a great judging sheet but it's only going to be as good as a person you have utilizing it. I think every competition host could do a better job of preparing the judge to be accountable and fair. Because you know how it is, you get there at the last minute, and it starts, and it's nobody's fault. Don't know if that can be done in an email prior, to just, you've got the rubric, but this rubric is going to be handed over to real people who they're going to be trying to improve. A lot of them may just want to get a score and win, who knows. But there are teachers out there that want this rubric, and

want to have information that they can use to improve. I think that's something that can be addressed and in every show choir competition because we lose sight of what we say on that tape and what we write down on this score sheet is going to be handed over to students. And we need to be as positive as we can because it's just our livelihood, whether groups are prepared or not. You want everybody to go through the next level or to be as successful as they could possibly be.

I: One other topic that has come up a lot and it wasn't a part of my original questions but it keeps coming up. So, I just started asking everybody about their opinion. How do you feel about caption scoring and what I mean about that, and you've kind of touched on it, that you have a panel of judges and within that panel, you have judges who only score the vocal portion of the sheet and judges who only score the visual portion of the sheet? How do you feel about that?

W: Personally, when I've done that, I don't like it because it's like, okay, I'm only responsible for this, but in the classroom, I was responsible for all of it. To me, if you do that you're not using everybody's expertise. Some people, you're using too much of their expertise. Like if I do just vocal performance I maybe stick more on tone or diction or whatever but don't really connect that to what they're doing with their body. Visually, I think, I think you've wanted to judge the whole shebang, all to it. And because to me that's the only way you could be fair and give people things to improve on. If you just do the vocal part of it, like you said, Ron says "You put your head down" you're doing half a job.

I: I think that's interesting, because the people who have spoke positively or in favor of caption judging have been people who aren't necessarily in the classroom every day. And that's your background, as an educator, as a director, you don't get to separate it when you're teaching, you have to look at all to it and teach all of it. So, I think that that's interesting.

W: Yeah, yeah, and the vocal performance, I mean, I've said so many times, what you're doing choreographically, is phenomenal, but it's not allowing you to be successful vocally. And if I was just judging just the vocal part, I might not have said that or I might have... I said you've got eight parts, but you're singing 8 parts and are concentrating on that, you're not able to really sing through the choreography.



## **Appendix C.2: “Mr. Southward”**

(I = Interviewer; S = “Mr. Southward”)

I: So, I've got a script, and I'm just going to go by the script and I'm going to start reading and then there'll be some questions so thank you for taking the time to answer a few questions regarding your perceptions of the Tyson Rubric. This interview is being recorded, and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential, by continuing on, you are giving your consent to the interview and the publishing of the resulting data is that okay with you?

S: Yep.

I: Okay so there are 13 questions, some are very brief and some feel free to elaborate on. If I would like you to elaborate, I may give prompts but if you elaborate too... it's fine what answer however you feel sometimes yes or No is okay.

S: Okay.

I: Okay, could you tell me your current occupation?

S: So, I'm a show choir choreographer and then also an actor.

I: Okay, could you briefly describe the types of score sheets that you have used at show choir contest? Are the rubrics global score sheets, meaning there's one score for everything. Are they weighted are they not weighted? Just your general impression of score sheets, you have used when judging.

S: Sure. Think I've used almost all of them, I've used digital ones I've used state rubrics, like ISSMA and Indiana State, also used the one that you use in Mississippi? And then even, I've even used the one that has done by, I think it's ACDA for show choirs and I think it's maybe even jazz choirs, I think it's like the same rubric They use for both. So, I've done all of those so far.

I: Do you feel that in general, the ones that you have used that the total points available for scoring are accurate? Are they too wide, too shallow? Just about the total points available for you.

S: I think often times it's not descriptive enough, or doesn't value things enough meaning the one that's coming to mind is the Indiana State score sheet is one big, broad category for each thing and it really doesn't give someone a frame of reference of what to work on or where to go from. I don't feel like it's educational at all, like it's just a tool.

I: Okay, and then in your opinion, do you feel that evaluation sections should be weighted meaning that some portions of the total score are worth more than others?

S: I feel that certain sections should be, meaning like, I think because it's show choir. I think the vocals should always have the highest point category. I don't know that I would say individual categories under each of those. If you had a vocal section choreo, section show band or what, however that sheet may be laid out. I don't know that each category under that should be weighted because I know certain people have different techniques or values based upon, I guess, certain categories, so where someone may feel like the most important thing would be technique. Someone else may disagree and think it should be something else, based upon age range or things of that nature, but I think that's probably part of your project and I think everyone is coming to the table with different biases and perspectives and desires. So I think definitely as far as what the emphasis should be,

should be weighted. But I have mixed feelings. I think maybe about the categories themselves, being weighted differently, I also think that can get confusing, too, when trying to think How scoring goes, because I'm such a math person because I've been on the receiving end of scores, so I like to think a seven out of 10 is a 70%. So, this means someone is average, where I know a lot of other people don't think that way. So, for me, if it's out of a 5, then I'm going Okay, what's a three out of five. I have to think mathematically. How does that interpret to the people who are... So, the score means something specific. So, I think sometimes the categories can get a little hard if it's not, if it's different for each one. If that makes sense.

I: Okay, you spoke just a second ago about bias could you speak to the problem of bias in choral adjudication?

S: Yeah, I think everyone comes to the table with something different, meaning, but yeah, I think the main thing I notice is some people all judge with a... I'll never go. Let's say it's out of 10 point someone will say, "Well I never judge below a 6. And then some people were like... Well, I always judge below a 6 or... I think everyone comes to the table with a completely different agenda I would say, So I... And so then, I think not everyone can be unified in the goal of what the adjudication would be for the day. So I think that's hard, and I do think to if we're looking at show for instance, I think we've got a certain part of the country who values storytelling and wants those elements in place in a show choir show or maybe someone else on the other end is just looking for just good singing, good dancing and just looking for technical things, and ways to improve those things. So, I think it's a matter of, there are people with artistic emphasis, and there are people with technique and performance emphasis, so I think those biases are also conflicting at times, so I think that's the ones I see most often. So I think people have different expectations and goals when they're judging. And maybe not on the exact same page when approaching a score sheet.

I: Okay, does the type of score sheet used have any impact on adjudicator bias?

S: I think it does, sometimes it does definitely for a choreography category, I think sometimes it can be more artistic or what's the word I'm looking for more design-based rather than execution based? And I think once again, are we judging the artistic merits of the creative staff or are we judging the execution and the performance that the students are giving on stage? But I think that depends on the score sheet, because I've seen many leaning towards... How is the creation and how was that done versus what are the students actually giving us on stage and so I think it depends, yeah.

I: Could you speak to the usability of the Tyson rubric from an adjudicators point of view?

S: What do you can go in more detail with what you mean by that?

I: I can... So even though in Mississippi, and we use the Tyson rubric you're scoring on via Carman scoring systems. The rubric itself is usually you're given a sheet of paper that contains the rubric. Do you feel that it's user-friendly from an adjudicator standpoint or have you used other things that perhaps are easier to use or is there a better way to use it or do you feel that it's fine the way it is?

S: Sure, okay, so I think the biggest thing I like about it is that it's consistent for Mississippi meaning whether you're judging on or receiving scores on it, you don't have to question week-to-week a different category that you mean it might be frustrating if

you're competing out of state and it's different, but I think the thing I like the most about it, is it's familiar at this point just because people are all they know that category is going to be on their every week, or they know how many points are in that category and every week, I think the consistency is like great. Where, in the Midwest, where I'm judging a lot every week, it's a different score sheet, different point total different everything, almost every week. So, I like that part about it. I do think, like I said earlier, I think it's hard with the weighted categories at times, just because I think when I go through after because I always review everything I do before I move on to the next group, sometimes I get nervous when I'm like... Oh wait, that says a five. Wait, I didn't mean to put a five there and I'm like, "Oh crap, that was out of five, okay, then I have to re... I think sometimes that's hard just, when re-evaluating how I judged just to make sure that I was doing it accurately what I was thinking. So, I think that can get confusing at times. I love that it's digital, I so love that it's digital because that makes it 10 times easier. You can just click and go. I do wish there was more of... I love the rubric, but I wish you could break it down just a little bit more just so I think I have an Ohio group that has a contest and they go over it in the morning and it's like a... We feel that this point means this or... And I think you basically have on the rubric, but maybe just making sure that everyone's on the same page.

I: So you mean that someone at that competition is doing a form of judge's training or orientation?

S: Yeah, yeah, I think just... Just that morning like meaning on this score sheet. We, view a 6 is this, is this view you know, what I mean? because some people might be like... Well, my five means something different than your five. Does it make sense?

I: Yes.

S: I think that would be it, that way, just I think, so it's consistent across the board.

I: Okay, and you've kind of talked about this a little bit, if you want to elaborate more, how does the Tyson rubric compare to other score sheets, you have used?

S: I think the biggest thing I love, but it's just consistent. I do wish at times there was more categories in the visual areas. I think sometimes a lot of the ones they see in the Midwest are a little bit more detail-oriented about certain things that they hope to achieve or that they're looking for in the visual category. But I think overall, I just like that it's consistent. I think that's my favorite thing about it because I... Everyone in the state use of it and it's familiar with it and knows what to expect.

I: Okay, and this is going backwards just a little bit. You touch... Touch on this briefly, what sort of orientation or training did you receive for the Tyson rubric prior to the contest that you judged? This could include receiving a copy of the rubric prior to the contest or did you not receive any orientation for the contest that you judged that used it this year?

S: I think basically what I got was I got the email from you I'm pretty sure that's said you're a judge for a contest in Mississippi. Here's the rubric please review it. And then I think it was... Get back with you, with any questions. And then we had a follow up with you, but that's what I remember the most. I don't... Which was great. But I don't think anyone at the specific contest, I was at really talked about, I think it was mainly coming from you.

I: Do you feel there are any benefits of the type of rubric for show choir adjudication?

S: I let me think... Yes, I'm just trying to formulate why. Yeah, I definitely, I think I wish in my perfect world, that everyone could use it so that there's not a change week-to-week. That's obviously a dream but yeah, I think I like that it's, I like that it's user friendly, definitely on the judge side, and I like that it's digital and I like that it's instant that you can almost get those right as on as awards are done for people, who need to review them and look at it. I do think it's very helpful. I do wish there was, Like I said, I think in the visual category, just more research and more scope of what visual people are looking for a lot of times or visual-based, people are looking for... because I do think it misses some scope there of ways that the students can improve technique and performance and execution, but in general... I think it's... I remember loving the vocal categories. I remember loving that blend, but I'm pretty sure if I remember this correctly, but blend and balance are two separate categories, yes, which I don't see all the time. And I love that there's a differentiation with those. So, in general, I really did like it.

I: Okay, did you detect any limitations of the rubric? So, I know that's really broad but...

S: Yeah, besides just wanting some more specifics in the visual category and just that some of the categories itself, were a different number, I think it was hard to... I think that was a limit for me it was, I was like, "Oh well, I feel like this category for instance, should be out of 10 because I think. It's just as important as something else. Okay, I think that was the limit for me at times. And trying to re-calibrate your brain to know what did that... 2.5 out of five equate to a number you know, what, I mean, because percentage in that or how does that work.

I: And my next question is, if you thought that the categories were appropriately weighted and if not, what weights might be more effective in giving appropriate balance of importance? So do you think that they should... All of those sub-categories that you're talking about, should all be worth the same amount or on the same scale, or how do you feel about how it's weighted?

S: Sure, that's a good question, I don't know. And part of me wonders if the category, maybe isn't worth as much, should it be on the sheet? This is me speaking aloud I...

I: No, yeah, this is great.

S: It should, it should have been. It is valuable or as important as other things. And also, I think I'm seeing a lot of times too, where I think both up north in the Midwest, where finals is a more comprehensive sheet or finals has then become a less comprehensive sheet where they're only looking for certain things like that, though. I know some people in the south do the fair thing where it's like an average. And ranking, et cetera. But I have seen that happen more where it's just very specific things is what they're trying to achieve and look for the evening. So, I don't know, I think that's a really good question. I feel like I don't have a good answer...

I: It's fine.

S: Okay, yeah, yeah, maybe just considering if it wasn't as important as other things, would it be worth putting on a score sheet and if it is on the score sheet could other people value it the same way as they or put the same importance into it as other things? Okay, if that makes sense?

I: It does. Any other thoughts about the rubric, itself?

S: No, I loved it, I like that you actually gave what your goal and hope was for each category. And each thing too, because I think that's also not gone through too often

before contests and actually get someone a frame of reference were to go from versus like I said. Though, I do feel most people probably don't read, which I'm the opposite. I feel like I have to read everything because I'm so detail-oriented, but I feel like some people may ignore that or just skim it and not actually soak in what you're sending the... But I like that you did that because I think it's an attempt for everyone to kind of judge in the same way to try to help these people with feedback, and somewhere to go from to make it better.

I: Are there other issues concerning show choir scoring that should be addressed in a rubric?

S: Can you go more detail about that, are you saying your scoresheet specifically or just in general?

I: In general, do you feel that there's something else that's not really currently in any rubric or score sheet that you feel should be added or addressed?

S: I think looking at it in just in general, at the specifics of what choreographers and local directors are teaching their students is making sure that those are on the score sheet. I keep going back to in my opinion, the worst score sheet I ever scored on is the Indiana state scoresheet. I think it's craziness it's almost like I don't know if you guys have solo and ensemble at Mississippi, I'm sure you probably do, but they have a one through five kind of thing, and it's like one means da da da... and two means something, but it's hard sometimes when somebody has a lot of things in one and a lot of things in two, but you have to put a certain point number to it, it's hard. So, just I think finding ways to have things broken down in categories that when a group is really good at one thing that can be rewarded for it and when a group really needs to improve and is struggling in a certain area that it's apparent that that needs to happen, so I guess rewarding people when they're doing something really well, making sure those categories are there to affirm them and to reward them for it, also giving them specifics of where they can improve if that makes sense.

I: It does. And the last question, just if you have any other thoughts about either the Tyson rubric, or show choir scoring in general, positive, negative, any closing thoughts that you would like included?

S: Sure, I think there's a big debate that I'm seeing right now with two different things. One is judges judging the whole sheet. And judges doing their certain area of expertise, that's a big one. I keep seeing that it's coming up right now. I also, there's another one that's kind of starting right now, where we've got... And it's kind of like on the d.l. But I'll share it because you and it's your thing where should people be eligible to judge if they don't have formal training or qualifications for that, meaning if someone doesn't have a degree or some kind of formal training, and industry experience should they be able to give feedback and comments on something if they don't really have any real world knowledge of that has been a big one. I'm seeing a lot, yeah. And then the third one I'm seeing is division and categories like what are we constituting as A large school or small school, or do we not go by school size... And do we go by how big the group is and if we go by how big the group is if they have alternates and swings does that count into that or do we go by the number of people that they actually put on stage? And then if they do that, is that fair? They've got 15 people in the wings to go in the moment and are using them as an advantage to change categories. I feel like all these things that keep seeing

they would be tossed around of how can we make it more fair on... But judging me more even, and more on a fair playing field. Yeah, I don't know if I have like specific feelings on all those things. I just see that topic, those topics debates coming up a lot recently.

I: Yeah, same I would really like your opinion if you have one on the first issue that you mentioned which was about having judges that specifically only judge one category, for example, someone with maybe vocal expertise and more of a vocal background is judging only the vocal portions and someone who has been trained in choreography and dance and... Is judging that. That was a lot of the feedback that I've gotten, and I was just wondering, what do you think that as a judge you would like to only focus on one area or do you think that it's fair for everyone to judge everything?

S: I think it depends on the person. I think that goes back to the qualifications and training thing. Anne Chapman comes to mind were obviously she's kind of dual-trained, or for me, mine was vocal degree with dance involved. So I think it just depends on the people you're bringing. Like Stephen Todd is another person I'm thinking of has a Master's in vocal performance but also studied musical theater, and has crazy dancing. So I think it depends on the people. I can see the benefits to it, of having... I've seen often up north. It's like all that. Anything that you see has been handled by a visual person, meaning design choreography, etcetera, was handled by a visual judge, and then anything you were hearing was done by a music vocal judge.

I've seen that pretty consistently, and then I've seen even to switch it up where to balance it out in finals everyone judged everything. They left their categories. But I know, I know on the receiving and as a choreographer I've had people like a vocal judge say port-a-bras with your feet. Well, that doesn't even exist, it you know that's saying there are certain things like that or I've had a choreography judge say if the veins in your neck are popping out, you're giving it the right intensity, which I'm totally against for vocal tension. So like it's hard because then you get some people who are making comments and giving feedback that are not helpful and are not educational or giving you a place to go. So, I see the benefit to both, but I think it's also I could see it also being feeling limiting for people like I mentioned, to have such extensive knowledge and training and both sides like I think it's helpful for people to judge the areas that they're most qualified in and I guess if the way you could appease people would be to do that final thing where once you make finals, then you could judge all the categories, but once again, people might get upset with that because it could completely change the ranking. If someone who is a choreography judge is now judging vocals and completely disagrees with the local judge. But yeah, I do see the benefits for sure, of letting those people judge things they know best, because I do think that the limits... The kind of crazy comments, you can get from people. So yeah, so did that help.

### **Appendix C.3: “Ms. Alan”**

(I = Interviewer; A = “Ms. Alan”)

I: Regarding your perceptions, this interview is being recorded and will then be transcribed and coded, you will not be identified in the resulting data and all of your information will be confidential, by continuing on, you are giving your consent to the interview, and the publishing of the resulting data. Does this sound okay to you?

A: Yes.

I: Okay, the first question, could you tell me your current occupation?

A: I am a show choir choreographer and adjudicator and clinician.

I: Great, could you briefly describe the types of score sheets that you have used at show choir contests? For example, are the rubrics are they if... So are they weighted meaning that some categories receive more point weight, than others, or have you used global score sheets? And what I mean by that, or is it just... Here's a number from zero to 100, and you circle a number.

I: So, if you could just briefly describe those score sheets that you've used.

A: I've used a variety of score sheets... For the... When I've been in Mississippi and in the South, I've often used the rubric and then there have also been other place, I want to say a few times, FAME for a while was using a rubric system and then they changed because they started doing caption scoring, especially during their preliminary round. And so I was only doing visual scoring and what that means. I know they put more emphasis on the vocal score than the visual score, which I do think is important... And then there are other times where I know in Indiana, with their ISSMA scoring. But I believe they've just changed it this year. It was a totally different rubric situation. It's very much like taking the SAT. It was very, very difficult, but I loved it. But I do believe they changed theirs just in the last year to make it a little more in line with other competitions and just make it a little more user-friendly for all because I think they were finding that maybe some people weren't competing as much and using it because it was a little more complicated and it was... We had to circle things, and draw lines and it was really, really complicated, but I loved it because it made me... I think it wasn't as broad. They were really limited with what we were doing as visual adjudicators and vocal adjudicators. And I thought it was fun, because as someone coming in from the dance perspective who actually knows technical dance, I got to be very technical and that's what they wanted me to do. They wanted to say hey do these people know dance... Are they getting a good dance instruction or they just swapping around on stage? I enjoyed that because with vocals, there is a lot of technique that's being taught to the kids, but a lot of times you're just teaching 5-6-7-8 to the kids in the dancing and this allowed the groups that are actually getting some quality dance instruction to be rewarded and I like that a lot...

I: Great!

A: so that was kind of fun. Yeah, so there you have it. I've seen just about every time of score sheet there is.

I: which is exactly why I wanted your opinion. Okay, could you speak to the problem of bias in choral adjudication?

A: I think its kind of geographical because everybody has a different gear and I think everybody has a different idea of what good is. I find, as a person who travels a lot, and

judges all over everyone has a different idea of what the right sound, is and they emphasize different things. And so, I think that's what I find the most. That's what I think is hard because I think there's no wrong or right because if they're singing into it, they're singing well, just because one has a wobblier vibrato than another doesn't mean it's wrong. It's just that in some states, that's what's promoted and then in others it's not... So, in Iowa and Indiana, they want big, wobbly vibrato. And then kind of in Nebraska, they wanted to be in the middle and then in Illinois, not as much. They definitely want to adhere to the style of singing, and very much less vibrato. You know in the, in the South it's kind of in the middle... And so, it's just really interesting to find where you are geographically to know what's right.

I: Do you feel, in terms of personal bias do you feel that there are perhaps judges who struggle to separate their knowledge of a particular school group or director when judging?

A: Yes, I've seen it. I really have a problem with this and I try so hard myself not to do those gigs with that. and if... I'm so glad that I learned from some really smart people early on to not do that. I think that's why, I think that's why I work a lot. Just because I'm fair, I call it like I see it that day and I have a lot of friends all over the place, but they know that I will still love them on Sunday after the competition. But I do see that so often, and I just... The unfortunate thing is, we are at the point that there are so many different competitions on every given Saturday that the number of judges, to go around and put qualified judges can't be spread around enough. And so I do find that sometimes you do run into that, but one thing that I've found that, And I promote this a lot to schools that I go to, and a lot of the schools already do this, because they have seen it elsewhere, because they've competed at invitationals that do this, or the directors running the event do this, and I believe that FAME was the first, when Chad was there running it, and I know that part of America now does this. There is a judge's accountability contract, that we sign And I love it. And not only do we sign it that morning we talk about it, and we go over the bullet points and they say You will adjudicate what you see here and are exposed to today. And I find that so very important and that what you are evaluating today, is not based on reputation or director or choreographer, it is what those children are doing on that stage, today? Because it's not about the director or the choreographer.

I: Hey, and I lost you, can you hear me?

A: Oh, there you are.

I: Okay, good, I could hear what you said, I just wasn't sure if the recording picked it up, okay?

A: It's about what those kids are doing for those seventeen or whatever minutes on the stage. If it's not about choreographer reputations, director reputation. It's about the kids.

I: Kind of in the same line of questioning, do you feel that the type of score sheet used has any impact on adjudicator bias?

A: I think it can, I think they are good score sheets and there are bad score sheets for sure, because there are sometimes things that are on there, that can't control what's evaluated. There are things that I want to give points for and I can't, or things that I'm having to evaluate that have nothing to do with it and so, definitely because you're really limited to what that score sheet says... And I am truly a person, and this is one thing I've really learned over the last... I really try to just stay so true to that score sheet. and that's all



you're evaluating on... because no place on there does it say, "Do you like this? And I've done a lot of clinics and I've even done some for teachers, and I say you all spend so much time working on all of your team development and no place on that score sheet does it ask about your team development.

I: Leaving that topic for a moment. Could you speak specifically? The next few questions are about the Tyson rubric, and that's just what we are calling it for the purposes of this research, the rubric that is used in Mississippi now. First, I would like you to speak to the usability of the rubric from an adjudicators point of view.

A: I think it's really easy and I don't think it's too complicated, and I like how it's worded. I think anybody can understand it, which I really enjoy, and I like the fact that I'm not having to flip through the pages, because I've dealt with some... That are so complicated that you're always having to go back and check yourself, and I like that.

I: Okay, great.

A: and when you go do it, you're confident in what you're doing and you can stay consistent with it, which I like as well.

I: Okay, how does the Tyson rubric compare to other score sheets that you have used? You kind of touched on that, but if there are any other ways that you could compare it to other score sheets, you have used.

A: I think I can get comparable to most. The few things that there are only a couple captions that I think could provide a little gray area and they are controversial and it's just because everyone can pick their own interpretation and that's when you have the difficulty. When you're speaking vocally, if you don't have the score in front of you, you're not counting how blown up a chord is. and a lot of times, a director can fight and say "Well we're singing six-part music, we're singing 8-part music. But here's the thing, I don't see it and that score's not in front of me. And here's the thing too, there are a lot of times that some arranger, can be writing it 6 and 8- part, music for those kids, but they're not singing it, so it's truly not that difficult. And also, things are moving so quickly, in those shows we're not counting how many parts are being sung at a time. So, difficulty vocally always controversial, I feel. And also, for choreographically, there can be groups that are clean as a whistle, but it's easy choreography and then also there are groups out there who are dancing their little faces off... And it's hard, hard, hard. It's difficulty on both ends of the score sheet, sometimes is a challenge. And there are also some people who don't have any movement background who just think it's a big ole hodgepodge mess up there, and they don't realize how difficult that choreography is, who strictly have a vocal background. So that, those are the only two components that I think sometimes categorically can be hard and everybody talks about it.

I: What orientation or training did you receive for the Tyson Rubric, prior to the contest you judged and this could include receiving a copy of the rubric prior to the contest.

A: Every time I received the copy in advance via email, which I appreciated so much so I can familiar myself with that and then at least one of the contests, I think we discussed it ahead of time, and they made sure that we were all familiar with it so that we weren't lost, we got there, which I appreciated so much because it's a situation you just can't be thrown in and be expected to know what's going on. because I know that I judged at least one of the events with someone who was unfamiliar with it, and so it was good to have that little

moment. And I think you need to, for people who is definitely out of state people and people who are unfamiliar with the system they have to have at least the sort of training. I think it's just beneficial. And I think for all the judges as well. I know at least one of the events we kind of sat down and said, "Okay this is base line for people and just discuss some things just to make sure we were all kind of on the same page. And so, to define hot button words and things because I think we all felt that we should be at least on the same playing field as well. But that's only fair to the kids.

I: So, okay, are there any benefits of the Tyson Rubric for show choir adjudication?

A: Oh, absolutely, because like I just said, I think it just keeps the judges more consistent and I'm certain for scoring, it keeps the judging the numbers more consistent because I've been in many situations where one judge can single-handedly call the event, because their numbers are skewed, and this is allowing, and granted everyone has their own kind of scoring and their own numbers. For me, I like a seven, I'm kind of the Russian judge. I... You have to really impress me to get a 9, and that's just how I go. But using that system, it makes me spread my numbers out a little bit, and I think that's good especially for me because I've had situations where I've had directors or contests come and say girl, if you want your opinion to be heard and validated, here today, you need to open up your numbers a little bit more or else so-and-so judge is going to call this contest today. And so, I think because some judges can do that because their point spread is so wide. And so, I think using that rubric, it really does help. And it also allows those judges who are kind of crazy to stay consistent with themselves. But I always find that I'm very consistent with myself in my judging. I don't have a tendency to judge on Mars very often.

I: No, you don't... You not make... And that's why you work so much.

A: exactly, and I think that's why though, having a system like that, sometimes it's really good, so it doesn't let some judge just go berserk during the day.

I: Did you detect any limitations of the rubric itself?

A: Oh, I don't think so at all. I didn't feel like I was restricted at all.

I: Okay, great. Did you think that the categories of the Tyson Rubric were appropriately weighted, and this can... You can speak to the larger categories. So, just as a refresher, the vocal categories do receive more points than the visual or show design and then also within that, there are certain sub-categories, that receive more points than others. For example, for example, execution of choreography receives more points than costumes. So, you can talk about specifics or overall however, if you feel that they were appropriately weighted and if not, what weights might be more effective in giving appropriate balance?

A: can you give me specific number values cause that's the one thing I do remember.

I: Yeah, I can let me, I don't have the rubric in front of me, but I can... we did change it a little bit this summer, so I just want to make sure that I pull up the correct one. But I want to look at the one that is in my dissertation. Okay, this is the old one. Okay, so for example, under vocal performance - tone, quality, intonation, vocal blend and balance and diction all received 10 points...

A: Which they should.

I: Dynamics, phrasing, receive... dynamics and phrasing each received seven points, so not as much. And then, projection, rhythmic accuracy, style, and difficulty received five points.

A: okay, the only one that I felt that was hard in back seven. Seven is weird. Tens and fives are easy. Seven, seven, is the one that I remember being weird, because your brain isn't used to going there because it's easy to think of mid-range numbers when you're on a five or a 10-point scale. But seven was hard. I remember that.

I: Under visual performance - execution and facial communication, each received 10 points. Intensity, staging, and transitions received seven points and style, costumes, and stage conduct each receives five points.

A: For me, transitions is huge and that should maybe, and actually style because the style isn't the kids' issue, because that's kind of a choreographer choice, but transitions that is totally... I think that could be weighted a little bit more heavily, because that that is totally something that kids and teachers can be in charge of. I think that the student can be... cause so many choreographic things are out of the kids' control, like costuming choices. I think if you can award them for things that are in their control, those are the ones that definitely needs to receive more points. Does that make sense?

I: It does totally make sense.

A: Sometimes, like me as a choreographer, I always explain to my kids, I'm going to do everything in my power to make sure that you're receiving points, not getting points taken away because this is something that's not your fault, so it's all is me, it's my choice, or it's the teacher's choice and you're getting evaluated on that. So, I have to do my darndest to make sure that you're going to be evaluated fairly and as well as you can on this. That's not your choice. So, like a transition. Those kids... And those teachers have to work and work and work to make sure that those transitions are as smooth as possible, and I think transitions can make or break the show.

I: I totally agree, I want to say that maybe that's something that we changed over the summer, but I can't remember.

A: Yeah, that because as that transitions are a nightmare.

I: I think that I can't remember the other one because I'm so in the old one right now. Yeah, but that's one of the things. Okay, just a couple of more questions, are there any other thoughts about this rubric specifically?

A: Oh, I think it's a really great thing and I think it's so nice that you're trying to streamline it for your whole state. For everybody to use it because it's just going to make... I think educationally, it's going to make the kids understand it, and the teachers understand it. Probably for me is it just what makes me weird because I was a teacher first, and I always kind of come from the educational perspective. Hopefully, those teachers are sharing the information with their kids... and just not keeping for themselves because then the kids aren't learning week to week.

I: Exactly and that is my hope as well. Are there any other issues concerning show choir scoring in general, that should be addressed in a rubric?

A: I don't gosh, I mean...

I: There may not be anything, some people had ideas and others have not, but anything that you think that is not currently being controlled within a rubric or score sheet that should be.

A: If I think of something, I will let you know.

I: Yeah, you can email it, and I can add it.

A: So, absolutely, as you know it is such an individual and personalized thing. I re-think daily, even on Facebook today, I've read something about how can people judge hair and make-up and people who place value on that. They shouldn't be judging. And I was like, "Well take me off the panel, because I mean, because they said, "Oh should girls have the same hair and make-up but I'm like, this is a... This is a visual medium friend, you have to place value on that if you're going to spend the time and money to put all your kids in the same costumes, the same shoes, so absolutely you have to make them look good, I mean, it just blows my mind. No, I'm not saying that they have to be presentable. We have to take a shower that morning, and brush our teeth. But right, it just blows my mind and then people... I think you're doing a great job.

I: Well thanks, and then the final...

A: or people will find value.

I: I hope so.

A: I just think it keeps people honest and I think there needs to be more of that. And the only thing I wish as well and this is what I found a lot of even down there, especially so, so, so, so much table talk so much.

I: What, say that again?

A: table talk. And that kind of blew my mind. I was like, "Oh no friends, no no,"

I: Yeah it's very interesting, and I wish that I... And maybe that's further research for me once this is done, is to focus more regionally. This is what's done in the South, and this is what's done in the Midwest.

#### **Appendix C.4: “Ms. Bruser”**

(I = Interviewer; B = “Ms. Bruser”)

I: This interview is being recorded and will then be transcribed and coded, you will not be identified in the resulting data and all of your information will be confidential. By continuing on, you are giving your consent to the interview, and the publishing of the resulting data. Is that okay with you?

B: Yes.

I: Okay, if you could tell me your current occupation.

B: I am chair of the music department and instructor at Jones College.

I: Could you briefly describe the types of score sheets that you have used when judging show choir contests. For example, are they rubrics or have you used perhaps global score sheets? And what I mean by that is, there's a number from zero to whatever and you just circle a number with no qualifications. If you could just talk about that.

B: I have not judged, very much in the show choir realm, so I'll admit that from the get-go. And the one other time that I did it, that I can remember was quite some time ago, and I think it was a global sheet, a little bit more open-ended, not nearly as comprehensive and detailed, as what the rubric was that I used when I judged in Purvis last year. That one, we used your rubric. And that's the only time I can remember off hand that I've judged.

I: Okay, what about for choral adjudication in general?

B: Never ever asks me to do it. Just to give you a little background. I moved here from Texas and was involved in TCDA, TMEA, and all of those conferences there. Moved here, attended a couple of ACDA things and some regional events, and I'll admit my own snobbish-ness because I looked at all of it and I went... Okay, well, this is not really worth it. It has grown tremendously in the 20 years that I've been in Mississippi, but I was also involved for many years in church music, which had very limited resources in any of those capacities, and so I just haven't really been in the loop and when I was hired at Jones, I was hired alongside Mr. Dunlap. And so when he ended up being the choir director, I ended up doing more of the musical theater and then I took on doing more and more private voice so I've done more in that realm that I have in the choral realm. I just never was part of the in-crowd, here in the state, so I don't get asked to do any of that.

I: What about any experience with NATS or anything vocal that has been scored?

B: I have done some of that and I've at least been witness to it on both ends, both from a competitor, and from a teacher standpoint. So yeah, some of that has... And even some of the solo comps I think some of the places where I've seen that, there's actually more resource there and a better rubric for judging solo stuff than there is for judging group.

I: Right, right, okay.

B: which seems odd that you would not have it in both places...

I: But, it does seem odd, but the more I'm finding out is that there is no rhyme reason really for any sort of adjudication that's going on right now in the vocal world at least. I think bands kind of have their act together. Bands of America, Drum Corps, Winter Guard, but chorally, it's just all over the place. Okay, in your experience or your opinion could you speak to the problem of bias in choral adjudication?

B: Sure, in my experience, both on individual vocal basis and choral basis, there tends to be a bias on the part of the adjudicator because of whatever style they teach and prefer. So if you have an instructor for example, who holds up Anton Armstrong as their model of a good choral sound, and he wants everything very rounded and hooty like the St. Olaf sort of model versus someone who loves the Cambridge sound, you're going to get very different comments because you're going to have the bias of that adjudicator. I think there are some things that escape that bias. So if you, for example, you can talk about intonation and either it's in tune, it's out of tune and that's going to be a little bit easier to identify and codify in your notes or in a score, but if you start getting into the opinions of how do you like your vowels shaped in your choir, if you're judging a choir that has a director that differs from you, then you're probably going to score them down.

I: Do you think that the type of score sheet used has any impact on that bias?

B: It might, I think it might help to eliminate it if what you're being asked to score is very specific, then I think that you can possibly tone some of that down. I think that would be a good place to start. Like I said, if you can pinpoint exactly what you want that adjudicator to score and not leave it open-ended, then you're more likely to get more accurate scoring.

I: We're referring to the rubric used in Mississippi as the Tyson rubric. Okay, so could you speak to the usability of the Tyson of rubric, from an adjudicator's point of view?

B: To be perfectly frank, I thought it was a bit cumbersome. It's very thorough, but it's a lot to have to do and try to use in a show choir performance that's 18 to 20 minutes. And I was not familiar with it going in. Had I had it ahead of time or had I been more familiar with it? I think with more and more use of it, it's going to become easier. But it wasn't something I was accustomed to, so I found myself having to leaf through... okay, what exactly am I scoring and what's going to give it that three versus that five or whatever the numbers were. And so it was a little difficult when I was under a time crunch. Very thorough, very specific, those things I really liked, but trying to use it when I didn't have much time and wasn't familiar was difficult.

I: You've kind of already covered this, but I'm going to ask anyway. How does the Tyson rubric compare to other score sheets you've used?

B: I thought it was very thorough and very well-done. The one thing... and amongst the other judges there we did discuss this?

the one thing that I didn't like that I think can lend to some confusion, and I'm not sure... I know there are ways to do it, I'm not great with statistical measurements, because that wasn't any of my area of expertise, but there are different numerical values for different things that you're judging based on that rubric. So, some things maybe on a seven-point scale, versus a 10-point scale, I think that's confusing to an adjudicator and, again, something that under a time crunch is not as effective for them to be able to use it quickly. I think there could be a way if there were items in there that you think needed to be, as a judge you would like to see weighted more, if intonation as an example, is one that you just think should be weighted 20% versus 10% for some of the other things in that particular category, there's a way to weight that afterwards rather than having the scale of the numerical values different.

I: What orientation or training did you receive for the Tyson rubric prior to the contest you judged? This could include receiving a copy of the rubric prior to the contest.

B: When we got there, we were given copies of the rubric, and shown the software system that we would be using on the tablets and we were given time to read over everything and to get familiar with it and sort of shown how the rubric we were given in print would translate to putting it into the tablet that we had. So, we were given time to read over everything, and it was briefly explained to us. But then it just takes a little bit to go through it just to know, and get familiar with it. So, it wasn't that any particular training was needed, we certainly understood it because it was very clear, it just took us a few minutes to sort of read through, and then understand how that was going to be used in that particular system. Because I know not everybody's using tablets or electronic devices, I'm sure across the state yet, but they were and so that made it a little bit faster and easier to assimilate it.

I: Okay, are there any benefits of the Tyson rubric for show choir adjudication?

B: Yes I think the Tyson rubric is a great tool. I think any tool that brings any kind of standardization is a good thing because I do think that judging varies wildly, particularly in show choir, which it's just such its own animal. And I'm not familiar enough to know what systems are used elsewhere because I know there's just a few pockets in the country where it's a big deal. You've got your Midwest schools, and then you've got, I think, some Florida schools. And so I don't know what systems are used there which might be interesting and that might have been part of your research, but anything that brings... Well, let me back up. If we think about it in terms of the fact that it is something that is done competitively there's no other arena where you do competitions where you don't know what the rules are or you don't know what you're being measured on. That's the thing, that's the definition in the craziness. That we... Oh, you didn't stand on your head? Well, I didn't know that was part of it. So in an athletic contest, anything like that, you just think of the extensive amount of rules and things that you have to go by that you know, if you incorporate these elements, it either changes the level of difficulty or there are so many things like that that I think in many other extra-curriculars are much more well-defined and show choir is way open-ended and it's become this, sadly, a game of one-upmanship rather than musicianship. So let me see if I can just incorporate four costume changes, instead of two or three, and that's going to make my show better. Well, that's, pardon me, that's ridiculous, and assine. That's not what this is supposed to be about. So with a good rubric and with a system in place that says these are the things and the qualifications you're going to be judged on, then you've got... you've leveled the playing field. UIL competition, which lord knows more of our schools around here should be doing this, if you're going into site reading, you know what you're up against and you know what your students are going to be expected to do because that has already a rigor in place and a system in place. And show choir somehow has just operated outside the bounds for a very long time, in my opinion. This is as a parent and as a musician. I sat, I've watched it for years and you go to different places and you just feel like you're playing a different game with different rules, and that shouldn't be the case. So, I find any sort of rubric that can be universally adopted, I think will be incredibly helpful and I think would be a great relief to instructors as well.

I: Great, did you detect any limitations of the Tyson rubric?

B: Not that I recall, again, I thought it was very thorough and I thought that it covered in various sections some good detail. And judges in a competition are always also going to have the option to give comments and to speak to the students as well as their instructor or the band because sometimes there are things that need to be said to the musicians that are playing for you as well, so you always have that in addition to it. But again, at least with the rubric you've got some good ground rules.

I: Okay, did you think, and you touched on this briefly, did you think that the categories of the Tyson rubric were appropriately weighted and if not, what weights might be more effective in giving appropriate balance of importance?

B: I don't know if I feel qualified to comment on that. I know this is where my bias could easily shine through. There are limits to the physical elements of dance that can be incorporated in show choir. You were limited by risers... were limited in some ways. So, I would love to see musical elements valued more than movement. I do think visuals. I do think overall. I love a through- story, I love music that is connected and thematic. I think that shows forethought on the part of the instructor, and I think it shows students a connection between music and how even things of different genres can be used together in different ways. So, I would always err on wanting to see musical things weighted more heavily than movement. But that's my own bias.

I: Yeah, are there any other thoughts concerning the use of the Tyson rubric itself before we move on?

B: I don't think so.

I: Okay, are there other issues concerning show choir scoring that should be addressed in a rubric scoring system or scoresheet that maybe aren't currently addressed?

B: I'm trying to think back through. I can't think of anything. And again, if there were things that any particular adjudicator wanted to address, they would certainly have that freedom to do so in comments.

I: Okay, and then finally, if you have any other thoughts that you might have about the Tyson rubric or show choir scoring in general, anything else that I didn't ask specifically, that you would like to share?

B: I don't think so.



### **Appendix C.5: “Mr. Field”**

(I = Interviewer; F = “Mr. Field”)

I: Thank you for taking the time to answer a few questions regarding your perceptions of the Tyson rubric. This interview is being recorded and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential. By continuing on, you are giving your consent to the interview and the publishing of the resulting data is that okay with you?

F: Yes, it is.

I: Thank you, so okay, so could you please tell me your current occupation?

F: I am a vocal teacher at Waukee Community Schools in Waukee, Iowa.

I: Could you briefly describe the types of score sheets that you have used at show choir contests? Are they rubrics, are they global score sheets, meaning you have some sort of numerical system from one to whatever number and you just circle a number. Are they weighted? If you could just briefly describe score sheets that you've used before.

F: Sure, I've used a wide variety of score sheets. Most of them have a numerical where most of the time it's based on my own personal opinion, and value system determine where I'm pretty much left up to my own economy to create where my five out of ten is or that sort of thing. Occasionally, there have... I think maybe one or two different contests I've been to offer a "This is what we're talking about, explain what that category is about... but they still can only... You're up to your own type of range/interpretation. I've used some that are a lot of writing, but I think primarily they're all numeric of some kind.

I: Okay, just going on that experience, what do you think about the total points available for scoring? Are they too wide, too shallow to provide adequate or fair judging or you feel that they are adequate?

F: You mean like total points for the whole sheet?

I: Yes, for the whole sheet.

F: I don't know, I... It really depends on the contest. Like if it's a 200-point score sheet the category of 20 that that is too... it depends.

I: Okay, do you feel that evaluation sections need to be weighted? And I'm talking about the larger evaluation sections, so vocal, visual, show performance. Do you feel that those should be weighted or should they all be worth the same amount of points?

F: Oftentimes, the vocal score... they may have more categories of the points in them, so I think they can take care of the weighting that way. With 60% of the points are vocal categories, maybe 30% so and then 10% in overall or something to that where they have a kind of sort of on a weight system by just loading what categories of that faction right?

I: Okay, could you speak to the problem of bias in choral adjudication?

F: Yes, I think there are a lot of... Well, I just, in vocals, the vocal part of the score you mean?

I: I think that I'm looking more for... Do you feel that there is a problem or with choral adjudication in general, meaning, bringing your personal...

F: yeah, I think there's lots of personal bias, I think I left to our demise, that don't have any kind of feedback parameters or category parameters that personal bias fall indirectly those concepts that do offer some sort of guidance or rubric, that those provide much more defined... to help unhook your own bias to align yourself with the provided

judgments. I think there are great deal of people, and it can even be regionally, depending on where the judges are from, what they feel is good quality vocal production and what isn't... or the preconceived notion, his group won last week so therefore they shall score high this week. Or I don't like that director, so therefore their kids are not going to sing well, I just don't like them.

I: Okay, and you spoke briefly to this already, but the next question is, does the type of score sheet, used have any impact on adjudicator bias?

F: Absolutely, I think score sheet that are just handed to me to say circle what you think doesn't give you any direction. And I think it definitely impacts the bias of it if there's not any parameters then I think, bias runs rampant.

I: Okay, for the purposes of this interview, and for my dissertation the rubric that's used at Mississippi contest is being called The Tyson rubric. So, some of the questions that are coming up refer to the Tyson rubric, and that's just the one that's used in Mississippi. Could you speak to the usability of the Tyson Rubric from an adjudicator's point of view?

F: I think it helps level the playing field. I think it really keeps judges focused and honest if, but they also have to be of the personality that they are going to follow that there's been... We all know judges that you can give them explicit information, but they're still going to circle whatever points they want, for their own bias. The judges that are true and honest will align themselves with a rubric and the ones that aren't are still going to be biased.

I: How does the Tyson rubric compare to other score sheets that you have used?

F: I found myself referring to it a great deal because it kept me more on target. If you have a long judging day, a group that happens at 8:30 in the morning, it helped me keep honest and true to the same kind of group that I saw, maybe at 4, p.m.

I: What orientation or training did you receive for the Tyson rubric prior to the contest or contests you judged that used it? This could also include receiving a copy of the rubric prior to the contest.

F: Yeah, if I remember correctly, I received a copy of the rubric ahead of time and had a time to look it over. I believe there was a set of some kind of instructions in that email that just walked it through a little bit, just important information about it. And then there was a on-site judges meeting that went over again. And to make sure the panel had the same interpretation of the rubric to just have a really good understanding. So, it was a more unified kind of mind set prior to using it.

I: Okay, are there any benefits of the Tyson Rubric for show choir adjudication?

F: Are there benefits of that rubric? Yeah, absolutely, I think it's tremendous. I think it does help -it just keeps people as judges more true to themselves, and I think it's also from two angles. I think one as a judge it helps me a great deal to keep me true and honest and away from bias. As a competing group, as a director of the competing group, I can now know what that director or that adjudicator's 5 out of 10 means, especially because they know like... "Oh well, this will know that I was lacking in X-Y-Z". So, I think it helps clearly there, and it's both ends. It just makes things more true and more honest and more... and have stronger feedback to help directors and students and just the expectation. It's almost like if you're going into a contest and you know they're going to use the Tyson Rubric you know, basically the answers to the test. And so if you're going in and you know that you're going to be evaluated on the following things with these

kind of caveats with different rubric, you're going to better prepare your students obviously there'd be some directors that won't give any thought or care behind that type of thing, but it can be very... I think it's really good for the classroom and for rehearsals for students about how to aspire to be a five out of 5 instead of being a 4 out of 5. Of course, there will be personal interpretation at every level of director. What is a five out of 5, and how they push that a little bit, but there is definite benefit to have it because it's an artifact that you can compare and stretch to.

I: Okay, did you detect any limitations of the Tyson rubric?

F: Not that I recall. I think it sometimes when I was vacillating and getting my mind especially getting to that 3:00 hour of judging it really helped me settle in and just determine where those points should fall because of just being true and to the types of statements in the rubric, and where they align and where the points fell.

I: Okay, did you think that the categories of the Tyson rubric were appropriately weighted? And if not, what weights might be more effective in giving appropriate balance?

F: If I remember right, I think they were all really pretty good. It seemed like there were some that were smaller point value, yes but I wonder if they would get lost in the ratio of points, if their weight would hold as true. But I don't know, I think to really fully feel that I would have to be a group that competed with that, or assess a day's worth of judging. If I looked at all the point values and read out just analyze the data to really feeling, fully feel if that really is the case, or if it's just a wonderment that I just... It's just a question that I didn't know if that had much impact.

I: Okay, do you have any other thoughts concerning the use of the Tyson rubric?

F: Not really, I just wonder though, if it will be standardized rubric, and does that make the competition stage, does it become too sterile? If it's a unified one that everybody uses, will it make every contest, like, will it sterilize what the ebb and flow will be. On the other hand, will that be a... is it good to level the playing field because of historic ways that scoring has happened in the show choir world, not only in the south and the Midwest and on and on. Does it help rectify some of that to keep it tighter? I guess a third thought is, is that, is this going to be a springboard for other contests to do something similar to give validity to their existing score sheets, or their kind of contest, I don't know.

I: Okay, are there other issues concerning show choir scoring that should be addressed in a rubric?

F: I don't think so. I think the Tyson Rubric really has... There's a lot of comprehensive things in there, and I think if I remember, I don't recall there being like... I really wish there would be blah, blah, blah in there.

I: Are there any thoughts you have about either the rubric or show choir scoring in general, either positive or negative?

F: No, I think I maybe covered it in the previous answers... the wonderment of what a standardized thing, the pluses and minuses or the questions or the maybes that can come out of that.

## **Appendix C.6: “Mr. Highland”**

(I = Interviewer; H: = “Mr. Highland”)

I: Okay, thank you for taking the time to answer a few questions regarding your perceptions of the Tyson rubric. This interview is being recorded and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential, by continuing on you are giving your consent to the interview and publishing of the resulting data. Are you okay with that?

H: Yeah.

I: okay, so the first question: what is your current occupation?

H: Well, current occupation... I started working for Loreal Paris in their active cosmetic division. I worked for the number one medically dispensed skin care division in the United States called Skinceuticals, and I was recruited by my president on a trip to Cabo, actually, and she got to know me the weekend she goes, I love you, I think you'll kill it in my company, do a great job, and so I waited a year for us to find the territory, and I woke up in April of this past year and she said, "Hey the Kansas City territory job is open, do you want it"? Basically, the next week, I was hired. And so, along with working at Loreal, I still freelancing. I'm still adjudicating this year, I'm still choreographing for colleges and high schools show choir groups.

I: Great, could you briefly describe the types of score sheets that you have used as an adjudicator at show choir contests? Are they rubrics, are they global score sheets, and what I mean by that is there's just a list of numbers like zero to 10 or zero or whatever, and you circle a number. Are they weighted, meaning do some categories of points received more than others? But if you could just kind of describe the different types of score sheets you've used.

H: I've used eJudicator which is a zero through 10, and I believe you can do .5. So, it's more like a Likert scale. I've used consensus ordinal ranking. Let's see what other... I can't remember, I had used Carmen another form of the scoring sheet. I've also used ISSMA, which is the Indiana scoresheet. And that one, it's really specific, because get the one through five, and the comment boxes that you have to justify why you gave the number 1-5, which I think it's a little bit more educational, but just looking at numbers guessing, "Well why was it a great amount of... For five? Those are the different types of stores, so just to recap, zero 1-10, then the ISSMA sheet 1-5, ordinals, which is scoring throughout the day and then switching to ranking in the night round, so I'm familiar with all different types of scoring.

I: Okay, what do you think about the total points available? Is it too wide or too shallow? Just in general, do you feel like some score sheets need more points or there's too much or do you have an opinion about that?

H: Well, I think the sense we're conditioned from first grade all the way to high school and even into college, I think that background educationally is to start with 100, start with the most points, and take away based off merit, meaning you're starting out with a perfect score before the first downbeat. And as we go through this, if I do things that could be improved upon or mistakes, then I deduct, as opposed to starting at zero and giving you points, as you go along. I think giving and starting at that optimistic 100 points, 10 points, whatever it is, I don't think the numbers matter, honestly whether it's one through 10 or

one through 100. I think what matters is the specific rubric that you're actually looking for? I kind of like the one through 10, because it's easier to justify 10 numbers than is 100 numbers in... If someone gives you a 48, you're like, "Oh well, that's a really random... Almost smack dab in the middle of one through 100. But if someone gives you a four or a five, it's like, Okay, I feel that I can see that better in my brain that that was about an average choice. But I think it comes down to rubric, rather than the numbers.

I: Okay, could you speak to the problem of bias in choral adjudication?

H: Yeah, so I think bias is inherent in the human condition. So, the things we're familiar with, the things that we know we tend to feel comfortable with... We tend to gravitate to things that we are familiar with, and we shy away from new things. Although the new things are the adventure things and the fun things, home is always with familiarity. And I think that carries over to the bias of adjudication and we're connected now, more than ever, to social media, YouTube links, twitter, you know, adjudicators are in the trenches a lot of time working as clinicians, working as choreographers, and I feel like their familiarity with some groups and some adjudicators and there could be that human condition of bias that is a part of the adjudication process. Because you have people who are familiar with a certain region who are working with certain groups in certain regions and so they're influenced by their experiences with those particular students, those particular choral directors. Sometimes they are influenced by the name of the group; the brand and the power of advertising and commercials as we're familiar with branding. And so sometimes a show choir's name says it all. And they must win because they're Clinton Attache or they're John Borroughs. Yeah, or I'm not familiar with this school. They must not be good. So, I think that is also part of the equation. So, I think it's important that people find adjudicators that really aren't part of a certain camp and they just do what they do because they love it. And at the end of the day, it is education, and it should be educational. And the minute you start putting bias and politics in it, you start cheating the educational side of things.

I: Okay, I agree. Does the type of score sheet, used to have any impact on adjudicator bias?

H: Again, I can't speak for other people. I have heard many times stories where you get a judge that's choral that's being brought into the show choir competition, and their scores are completely different than the rest of the panel. And I don't know that that is inherent in the scoresheet or if that's just bringing someone in from a completely different school of thought that is not aware of how to choose the members or maybe what they're looking for or the value of what they're looking for and they end up tipping the scale and really creating inconsistency with how the numbers shake out. But I think if you have a panel full of people are familiar with the genre and they're very confident about what they're looking for and the value of the numbers they're placing or what they're looking for, I think you can probably come out with a pretty accurate result with whatever scoresheet. I think it really comes down to qualification of who's sitting in that seat, whether you're doing caption judging or whether you're doing your comprehensive judging, I would get into that later, but I prefer caption judging. But I think it comes down to the adjudicator honestly, because the scoresheets have been around long enough that I feel like if you get a whole panel of actual qualified people who are comfortable judging show choir, you're going to get a pretty accurate result.

I: Okay, could you speak to the usability of the Tyson rubric from an adjudicator's point of view?

H: So, I'm going to have to get a refresher from you on the Tyson rubric...

I: So, I think what I'm asking is, even though we score using Carmen, so your scores itself are in Carmen, the rubric and the full descriptors of each category are on paper and it's like two or three sheets of paper, so they're not in the system that you're scoring on, they're really more of a reference. So, do you feel that that makes it more difficult to use or once you've used it and you've had an orientation and you're familiar with it, it doesn't really matter? So, I'm just asking, yeah, is it easy to use, or is it cumbersome?

H: I prefer to have the rubric with the actual software or the numbers that I'm thinking. I don't like them disassociated. I think that is also something that can subconsciously make you also disassociated with the numbers of the rubric if they're not together.

I: Okay, how does the Tyson rubric compare to other score sheets you have used?

H: What I remember of it? I think it's less elusive like some of the scoresheets. I think maybe different judges might interpret the rubrics differently. I like the way that it's pretty, I guess, specifically stated. It's well thought-out. I think that it's more direct rubric than maybe some of the other ones that I've used. I think you arrive at more the specific results and it kind of targets what your evaluation of what you're watching and listening to should be.

I: Okay, what orientation or training did you receive for the Tyson rubric prior to the contest you judged? This could include receiving a copy of the rubric prior to the contest.

H: You guys were so good. Well, you were so good about that. You sent it out in an email so we got it ahead of time and then before we were adjudicating we had a round table discussion about it and you opened the floor for any questions or maybe things that we didn't understand, but you would give the opportunity to actually talk about it and have a discussion and I find that helpful; in orients you to the experience. It brings everybody to the same page. So yeah, there was plenty of orientations to the type rubric.

I: Great, are there any benefits of the Tyson rubric for show choir adjudication?

H: Again, I think that you're arriving at something that could maybe be looked at being adopted around the country, regionally nationally, because I think you're trying to untangle a lot of, like you said, the biases, the misinterpretation of the rubric, what people should actually be watching and listening for instead of a not-targeted rubric. I like the specificity of what you're going for because I think it's educational, and I think it will, it will encourage less bias. I think it's aim is to, I guess, create some sort of agreement in the show choir world of... you know what, out of all the scoresheets this one seems to be the most fair when I go back to my students and we recap the competition and the score, I'm actually able to justify and educate based off the results, as opposed to having to just guess at what the results are.

I: Okay, did you detect any limitations of the Tyson rubric?

H: I don't think so, I really don't. I thought that I understood everything. I think I used it to its full capacity. Yeah, I don't think there was any limitations for me at all on the type of rubric.

I: Did you think that the categories were appropriately weighted? And if not, what weights might be more effective in giving appropriate balance of importance? And just a

little clarification because this question has been interpreted a couple of different ways. So, the vocal categories combined are worth more than visual and show design. So, you could speak to it in terms of that way overall, but also if you'll remember some categories received 10 points, some of the sub-categories, some received seven and some received five. So just however you want to answer that question but it's been interpreted both ways. I could have written that question a little bit better, probably.

H: Well, I think that the vocal part of it should be weighted a little bit more, a little more heavily than the choreography and show design parts, because it is show choir. I think that it's probably something that everyone can agree on. Again, this is just my opinion and that's not on the behalf of other people. Sure, but I think we had this conversation, and I'm not sure if it was your competition but I know things like appropriateness of style and difficulty are, I think those are two categories that can that can vary based off of what the adjudicator's idea of what that means. I think there's a couple of reasons for that. Obviously, the most experience people, I guess you could say educated as well, because I think experience begets education, I feel like style is such an important part of what makes what we do really fun.

I: Do you mean vocal style, or visual, like choreography style?

H: All of it. All of it. So, this is where when you get into caption judging, but this is where the caption judging really helps to mitigate not just bias, but you know it... I'm a choreographer and I have taken dance for many years of my life, and I explore different styles, and if there's a show choir that's doing five different songs from five different genres, I can really speak to that from a professional standpoint because I've got experience and knowing what I'm watching, knowing within the first few bars, oh is this a 1920's song? Is this swing? Is this going to be Latin? This is going to be something that I'm familiar with. If you study a lot of style then you could speak to it. Now, would I expect a choral director who sings, plays piano and is a really successful as a musician to understand some of the techniques in the style and dancing or choreography, not so much. That's not in their purview. So, I think going back to your question about was there any category that could be fit to mitigate some of the, I guess, inconsistencies in scoring? It would be that I would suggest caption judging and maybe flesh out a little bit more of what is expected out of difficulty, because difficulty is relative to the group's experience. The brand-new show choir, and the director's like, "you know what, I'm not going to be super ambitious with the first year or two because we're just getting our feet wet. Well, if an adjudicator is just like... Well, that was so easy. Well to those kids it's brand new, it wasn't easy for them. That's a very precarious category. Difficulty.

I: So that begs another question that's not on the script, but I would like to have your opinion, do you think that it would be helpful for adjudicators to know information like this is a new show choir, or this is an extra-curricular and we can only rehearse these days out of the week or after school. Or do you think that that imposes a level of bias or do you think that that makes you more reliable as adjudicator?

H: Well, I would say that harkens back to the theory of what you're trying to extract from that competition, and what I mean by this is are groups competing against themselves to be the best version of them? If that's the case, yeah, some of that history, some of that background information could be useful in assessing how far a group has come in two

years, in one year, because they're competing against themselves for achievement. Now, if you're coming from the school of thought they're competing with other groups, then maybe that background information can create bias. So, if you say, "Oh this group has only been around for nine months and they're super clean, and they sing really well and their clothes are pretty because you feel sympathy for them, or you feel like... Oh wow, you really come a long way. You might weight their score a little bit more because you're impressed that in months, they've done more than what comes apart to do in five years. And that's not what they're there for that day. But then again, it depends on what the goal is. If you're judging that show choir based off of their own merit and how far they've come and their potential... And if I just shakes out throughout the day, if you're judging each show choir based off that, and then your winner is chosen probably on to your last based off of adjudicating by the group's potential then I think background information is good. If it's competitive with group against group, I think background information might change a little bit of how we score.

I: Okay, good, do you have any thoughts that I haven't asked about the Tyson rubric itself.

H: No, I mean I asked you earlier what is the main goal of the Tyson rubric? What is it that of this work and involving all the people and crunching all the numbers, what is the end goal for the type of rubric? That would be my question, because I think if you arrive at that answer, it will prompt you to reach out to certain people to get to where you want to go. So, the rubric itself I felt really comfortable with it. And the only thing again is appropriateness of style and difficulty. I think those two things are the ones that are the most, I think they're interpreted differently I think in a lot of competitions. And that's just my opinion.

I: I think. Well, from what I've heard from other interviews the difficulty has come back up quite a bit. So, I think that that's one of the areas that we'll continue to revisit. Are there other issues concerning show choir scoring in general, that should be addressed in a rubric that aren't currently?

H: Well, I think before we were on the rubric, we were talking about the using of eJicator, I think that can go a long way for helping the accuracy and rooting out some of the biases that people can use within the score sheet. I would want to know more about that. I had like to know more about other people's or other professional's thoughts on how to choose adjudicators in different ways at different levels of competitions. For me, I think that is a new subject and I think it needs to be addressed. I think that's a really... I think that would be an interesting topic for most people who host competitions and who actually compete.

I: Okay, if you could tell me any other thoughts that you might have, either about the Tyson rubric or show or scoring in general, positive, negative, anything that you think would add value to the discussion that we haven't talked about or that I didn't ask.

H: I would like to see, I would like to see a unified sheet. I don't know if I'm saying unified as in every competition from your smallest level to your big national. Maybe show choir competitions should be ranked almost like conferences. Maybe where if you're in this top the score sheet is this, and then if you advance on, you are now using this score sheet and then is a national score sheet. And then if you go all the way, then you're at the conference level, because your scores have qualified you for such. And then



it makes the natural winner, I think, a true, it gives you more of a more accurate winner. So, for example if you have FAME competition and you have five of them and there's five different Grand Champions, and then they all go to the nationals, that's as close to what any composition has had done by offering a national competition. Then you get a true winner. But not a lot of show choirs... there's such a small percentage that are able to do that financially. So it would be nice to have a score sheet that kind of helps people do that, no matter what level they achieve. They know where they fit in because there's different tiers at different levels, whether it's the same score sheet, or whether it changes based off of your particular level. I think it can help people get a feel and a sense of where they are and what they're trying to achieve compared to some other groups.

I: Oh, that's a really neat idea. I like that. I'm kind of thinking about that for a second, I like that.

H: Now, yeah, and I think you're taking into account is this a new show choir, is this an old show choir, are people traveling across state lines, are they staying local? You know, you've got to wrap your head around then, what's the stakes for these groups? What's the stakes in the development and growth of their program? And I think having that sort of conference or tier level could be helpful because you're thinking about other things than just the singing, the dancing, and the musicianship. You're thinking about the genre as a whole and its progression and its growth in the country and furthering the art of show choir.

I: I want to go back just for a second, because I know what you mean when you say caption judging, but if you could speak to that a little bit, talk to me about what, that... your definition of caption judging and also why you feel that that may be the most valid and reliable way to choose adjudicators.

H: I think caption judging serves a couple of purposes. One, what's the point of finals if everyone's on a comprehensive sheet for the both day round and the night round? s You know, the scores may not change that much. If the group is holding back during the day because they know they stopped their competition and that they're going to make it in finals and they're going to really turn it on, is that really promoting good performance education? You're downplaying to the score sheet and you're playing to the process, you're not performing for the audience, you're now performing for a place in the finals or you're playing the system. If you have caption judging during the day where you have an expert in each field, you have a music arranger, you have an instrumentalist, you have a choreographer, you have a vocal technician. So, those experts are able to focus very clearly in their lane, so they can watch for the smallest of idiosyncrasies within the group and what they're trying to execute, they can really have a fine-tuned assessment of what's being presented. If you're comprehensive, you're going to lean heavily on what you're good at, and maybe it's not fair to the group that you are comprehensive judging, but you're a choir director and don't speak a lot about the choreography, because it's not what you're used to or that's not your expertise. On the side, you're a choreographer and a dancer, you're trying to speak to some of the choral things, you're probably going to get more choreography comments and adjudicate on that more than the choral part. So, if you were to go to caption judging during the day, and then a comprehensive sheet at night

there's a whole different ballgame. There's actually a reason to have finals now. I honestly think if it's comprehensive during the day and then at night there's really no reason for finals except to sell more hot dogs. It doesn't make any sense. The stakes aren't there and you're adding different people to the sheet. You're adding, oh that judge wasn't listening throughout the day, they were watching, and now they're listening at night, that's different. Or, same thing, this person was only listening during the day and now they're watching and listening at night. So, you get a different bag of numbers at night if you go to comprehensive when you were caption during the day. And I also love the idea of bringing the clinicians on at night, because they've only seen half the groups. So that also shakes things up a bit.

I: Yeah, well that has been one of the biggest feedback in items that I've gotten throughout the interviews. So much so that, I think... there's not time to implement it this year, but I think for next year we will probably go to that. So okay, is there anything else you'd like to state on the record before I stop the recording?

H: No, I just think you're doing great work.

### **Appendix C.7: “Mr. Silver”**

(I = Interviewer; S = “Mr. Silver”)

I: Okay, thank you for taking the time to answer a few questions regarding your perceptions of the Tyson rubric. This interview is being recorded and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential. By continuing on, you are giving your consent to the interview and publishing of the resulting data. Is that okay with you?

S: It definitely is.

I: Thank you, okay, if you could, could you briefly describe your current occupation and then follow that up with any sort of background or experience that you have in the show choir world as well?

S: Sure, happy to do that. Current occupation is a senior employee relations manager at Wells Fargo. So, I do that corporate thing during the week, I've been there for over 30 year. And my experience with show choir, is it started by being in groups like that all through high school and college but really got more into the high school competitive show choir stuff doing some choreography, and really basically an assistant director, for a group here in the Twin Cities and we competed in the Midwest and across the country for 13-14 years that I was just really with that one group. Did pretty well, and I think got my name out there and understood it got me exposure to really understanding what the competitive show choir is like, that whole environment. And so, it's just been really pretty extraordinary to start moving into more of just the adjudication stuff when I left that component of things. So, been judging for probably about 20 years or so, that's all throughout the Midwest and all across the country.

I: Okay, could you briefly describe the types of score sheets that you've used at show choir contests as an adjudicator. For example, are they rubrics, if so, are they weighted or is every category worth the same amount of points? Or have you used global score sheets? And what I mean by that is there are really no categories or descriptors. You circle a number either between zero to 10, or zero through 100, or if you could just kind of discuss the different types of score sheets that you're familiar with.

S: Sure, I'm taking some notes, so I don't forget a few things, but I would say all of the above. I think, for doing this, as long as I've done it, I feel like I've kind of seen sort of every possible sort of scoring approach. So, some of them are rubric- based and would be weighted in with very specific sorts of categories and descriptors like you were out running before. Where some are just incredibly generic, and you are left to your own devices to determine what's going on. There are some... as you know, there are some where you are only scoring a certain aspect, so if you're hired to be a vocal judge, or a visual judge... and then, you don't only really get to opine about what you've seen or heard, as relates to those and then they add those altogether.

I've also been in some where it is incredibly generic, as far as just a vocal and a visual where you really without a whole lot of categories within them, and you just are giving an overall score within that group. So, some electronic some on paper. It's been pretty much a little bit of everything.

I: Okay, could you speak to the problem of bias in choral adjudication?

S: Sure. I guess to me, and I get kind of opinionated about some of this, I think it can go either. I think of two things. I think of a very overall big picture of a bias, that can come into play or even ones that are just within a specific day or competition. So, giving you an example, and I don't know if you remember this from when I was at your place, last time. I'm kind of "that guy" that the end for example, at the end of the day time if there's a finals for example, I specifically asked to not know where other judges had their... Like, "Oh well, I have them in third or I had this" and I don't want to see a score sheet for that. And so, everyone's a little different. I think you can think of that in different ways, but I just would prefer to have that pure where you're not going to be swayed in any way, shape, or form even within a day because of how other people may have scored. So, I think that's sort of bias because I do think that that can be whether overt or a little bit more subtle can come into play at times. I think other things that I consider bias on a very big picture scale, is when a... and you asked me to give opinions so I'm going to give opinions. But I think at times, you hear things like... Oh, well that's so-and-so's group, that's one of so-and-so's group, and then usually that's referring to a choreographer. And LaDona, I've been a choreographer, so I'm not ripping on that particular function and what they contribute but I do feel like at times, that can be... Well that's one of so-and-so's and they're usually supposed to do pretty well. And you hear that kind of talk, and I just have always found that fascinating, with... That's one of so-and-so's group and as a choreographer and I always think... No, it's not, it's that director's group. So, I'm kind of particular about that, but I think that that can cause bias and I've seen that kind of happen.

I: Okay, does the type of score sheet, used to have any impact on adjudicator bias?

S: I don't know about the score sheet having bias. I'm trying to think about whether that might come into play. I don't know if it's bias necessarily but it may be kind of related to that but I do think having some specifics listed about what's within a category, can be really helpful because you may find someone who is scoring all categories visual and vocal who really is quite specialized in one of those two but not the other. And if there isn't, aren't those prompts that are there to really there to prompt and to think about things and go... Okay, I understand now what we're talking about, we're talking about balance or this or that, then it might be difficult for someone who maybe doesn't have the kind of background or experience in one of those two areas to be accurately scoring. So, I appreciate that detail in within it.

I: Okay, could you speak to the usability of the Tyson rubric from an adjudicator's point of view? Is it cumbersome, is it okay to use... is it a problem that it's not on the screen that you're actually scoring; that it's more of a reference if you could just talk to the usability?

S: Sure, well, I think sharing it well ahead of time with an expectation, that if you're going to be judging at this particular competition, this is really, what we're all attempting to adhere to, and we ask you to familiarize yourself as well as you can prior to coming here, I think that's always a positive, positive thing. So, I guess that's really the key thing that comes to my mind is that I think that just helps with that usability because you're more familiar with it, rather than trying to do it on the fly or and or I think a little mini-session before you start that day too. And having the director of the competition, I walk through some specific things and what his or her expectations are as well.

I: So, along those lines, what orientation or training did you receive for the rubric, prior to the contest that you judged this past year? This could include receiving a copy of the rubric prior to the contest.

S: Yeah, I do remember receiving that because I went to when, when I went, to Petal, so I remember receiving that ahead a time, it was sent out, I believe to everyone with a really great message and saying you've agreed to judge here in Mississippi, and this is what we would be looking for. So, I remember being very familiar with that and being able to look at it on the way there and then there was a brief orientation or revisiting of it once that judging panel came together. So, I think it worked out really quite well.

I: Great, how does the Tyson rubric specifically compare to other score sheets that you have used?

S: I don't feel that it was anything too dissimilar, I think that it just was, from what I recall about it is that, again, I just found it really self-explanatory. I will say to me, a component and this kind of ties to usability, I guess, as well, I do think that it can be difficult when there are a number of categories that have different point values assigned to them. And as you were kind of refreshing my memory earlier about some 10 points and seven points and five points, that piece can be... that can cause some confusion or just it's really having to stop and think. Okay, this is a seven for category, that six that I just gave is really high, but that's six I just gave on that 10.1... ooh. It's, I think, a little bit cumbersome at times when there is a lot of... And there are a lot of those. I personally just would prefer either they're all the same or you have just some really incredibly important ones for vocal and visual. One or two that are 20 and then the rest of them are all 10 or something. That's just a personal part of mine. And I think part of that, too, is I've seen judges, misunderstand that or forget about them. And you probably heard those stories too or experienced it where... Oh my gosh, so-and-so's been scoring that like it was a seven-point the whole time. And that's when I can create some issues at times.

I: Okay, are there any benefits of the Tyson Rubric for show choir adjudication?

S: benefits of just kind of the approach or the one in particular we're talking about... the Mississippi one?

I: The Mississippi one, the Mississippi one specifically.

S: Yeah, I really... again, I just think that the benefits of it are that it will lead towards consistency, and what's being reviewed by the judges, I just think that's a very critical thing, that it's all on the same, everyone's on the same page about things. I think there's great benefit.

I: Did you detect any limitations of the Tyson rubric?

S: And nothing that I would say would be a limitation. I think there is always... When you delve into any specific category there are different opinions about what weighting should be or how things are decided and nothing particularly jumped out at me. The things I will tell you, and maybe more detail than you're wanting, but I'm always intrigued by a category that is just difficulty alone. I'm getting really specific, LaDonna... but if a category is simply difficulty of literature for vocals or and/or difficulty of choreography, I struggle with that, especially if there isn't a very clear complementing category of how well was it executed in was it, quite frankly, within the level of achievability or attainability by the group. And where I'm going there is, you can have a group that has the most unbelievably difficult vocal stuff going on, but it's so far above

their abilities that it completely can derail an entire show. So that particular thing is one that always kind of tips at my eye. If there's something that's just difficulty without a complimenting to it, or a difficulty and execution kind of a balance sort of an approach. I am really, really particular.

I: Did you think that the categories of the Mississippi rubric were appropriately weighted, and if not, what weights might be more effective in giving appropriate balance of importance to particular categories?

S: Boy, I don't remember anything jumping out at me, that really stood out as a concern with that. And to be honest, I probably need to delve into that a little bit more detail again to really see what some of those were, but there was nothing that would jump out to me about that. I do think, back to my earlier point, the slight differentiation of what's the five-point scale, and a seven scale can sometimes be, it's very discretionary. But there was nothing that just come out and like... that just doesn't make sense.

I: Okay, do you have any other thoughts maybe that we haven't addressed or discussed concerning the rubric, itself, the Mississippi rubric itself?

S: Nothing that I can think of about that at all, no, and again, I think it's very clear and it's easy to use and understand.

I: Are there any other issues about show choir scoring in general, that maybe should be addressed in a rubric or a score sheet that aren't currently addressed?

S: And, well I guess for some reason that's what's coming to my mind, when it is about accompaniment and you referenced it, how things are broken down with the one that you use and that you come up with. I don't know that anybody's going to solve this right now but it's such a disparity depending on state about whether it's a group of adults or whether it's a group of students. And so, I know that that doesn't necessarily come into play when I come to Mississippi and it's pretty much every group are all adults so it's a level playing field, but I think that is kind of an interesting one as of how that works and how incredibly different that can be. And I do think with the recorded piece I appreciate that you have something specifically guided, about okay, if this is a recorded background, here's what really we're looking at for that so that it's not one group over another.

I: And then, finally, if you have any other thoughts about either the rubric, show choir scoring in general or things that we haven't discussed that you would like to include?

S: I think we've gotten to most everything to and I'm taking a look at my notes that I'd kind of put together before we talked...

Yeah, I just really what I do appreciate is that there's this component about overall effect, I think it's incredibly important because I will tell you one thing that... What I think is missing at times is a judge's ability to reward an appealing show. And I don't know how else to say that. You hear, "Oh my gosh, that group they know how to work the score sheet". You've heard that before, you probably heard that before. And to me, some of those shows, if it's simply based upon kind of those categories can sometimes be some that just miss as far as audience appeal, and I think, I get all philosophical here LaDona, but I also think what we're hoping to help educate the young performers about is how to relate to and connect with an audience, and so I just appreciate when there are categories like you have described, in that there is use there, that are about that overall effect, and it gives a chance for an adjudicator to say, "You're reaching... You're connecting with the audience or you're maybe not and I'm going to be able to reward you or not depending on

how well you do that", because I think sometimes that can get lost in that crazy show choir stuff that we do.

I: Sure, I have one more follow-up and it's not in the original script, but it's come up several times, and I would like your input. So how do you feel or do you have an opinion about caption judging? And what I mean by that is that you have a judge who's of vocal expertise that is just judging the vocal portion of the sheet, and then you have a choreography expert that is just judging the visual part of the sheet or do you feel that the entire panel should judge everything on the sheet? What is your opinion on that?

S: It's a great question because it is different. I'll tell you my first preference is that every judge would judge the same, and I think the really critical asterisk with that is finding the appropriately skilled judges to do that. And I think that can backfire if you don't. You've got, like I kind of referenced earlier, someone you may have someone on a panel who just is very fused from their talent and experience in only one of those categories. If you can find a panel really that have some balance in that background I would prefer that because, in my opinion, when I have the choreography right there, so wonderfully accented that vocal moment and I get it that if you still were in separate categories, you could reward that, but your mindset's really different. When I've been asked to judge one of those two you're so, I hate to say myopic but you kind of are, about what you're looking or listening to that you kind of lose sight of that overall thing. I think a second choice, and I've seen this a couple of times, I'm trying to even remember, LaDona, where I saw this or where this was part of it, where the panel was divided into a visual or vocal but they had, for example, a vocal judge with... I'm making this up, 10 very specific categories around vocal but they also then had two categories of visual overall show. So, it got two that were much more generic, whereas the visual judges would have can category in the visual but a couple around vocal, like a hybrid. And I kind of got that actually. I thought at least that gave that opinion to still be in their entirety of what they're reviewing.

I: Oh, that's really interesting.

S: Yeah it was kind of interesting doing it. It's been a while since I've seen that, but I have.

### **Appendix C.8: “Mr. Hayworth:**

(I = Interviewer; H = “Mr. Hayworth”)

I: Okay, thank you for taking the time to answer a few questions regarding your perceptions of the Tyson rubric. This interview is being recorded will then be transcribed and coded, you will not be identified in the resulting data and all of your information will be kept confidential. By continuing on, you're giving your consent to the interview, and the publishing of the resulting data is that okay with you?

H: yeah.

I: okay, if you could, could you tell me your current occupation.

H: Instructor of music and voice at Alabama State University.

I: Great, could you briefly describe the types of score sheets, you have used at show choir contests, meaning are they rubrics, if so are they weighted, or have you used perhaps global score sheets, meaning there really aren't descriptors or categories you're just given a number between zero and 10, or whatever, and you circle one of those numbers. If you could just kind of talk about what different types of scoresheets you've used.

H: Before, I've used the weighted one. I've used the zero to ten one, and there's another... Yeah, those are the only two I've used...

I: Okay.

H: Weighted and 0-10.

I: Do you believe that evaluation sections need to be weighted, meaning should vocals and visuals be worth the same amount of points or should they be weighted differently?

H: I think they should be weighted differently because it show choir; choir is more important and the vocal part is more important than the visuals. I don't think they should be equally weighted.

I: Okay, could you speak to the problem of bias in choral adjudication?

H: Of bias?

I: Yes, if you feel that there is a problem of bias in choral adjudication, if not, that's completely fine too.

H: I don't think so, not with working people that I judged with, for the places I've judged. I think everyone was pretty fair. Some people, there is a little bit of where people think that because a school might have a certain amount of singers, that they should do well, or a certain amount of people that they are building better because they have a reputation of doing well. But sometimes you can't win everything and it's just the person's opinion, personal opinion. So, it might not mesh with what your vibe is going for, or the performance that you're doing, or the theme they're going with... But, it subjective. It's not necessarily always going to be perfect.

I: Does the type of score sheet used to have any impact on bias?

H: no, not the latest one that you created. I think it does not at all because you have to back up your reasoning for it for giving the person that score, which make all of the sense. You can't just give them a six because you think they deserve a six.

I: Oh yeah, okay. Could you speak to the usability or ease of use of the Tyson rubric, from an adjudicators point of view?



H: It's the easiest that I ever think because explain difference between category, and explain exactly what we should look for as, in that specific moment and where we should, where we should, for the ensemble. It's the best that I have used so far.

I: How does the Tyson rubric compare to other score sheets you have used?

H: Other score sheets mainly were for percentages. It would say between 70 and 80, or 80 and 90, 90-100, they wouldn't go lower than that. You can mark an x, or the x marks the spot. So, where you think that percentage is. It's just a... It was just too subjective... There was no... No way to explain why you think this group should get 8.

I: Okay, what type of orientation or training did you receive for the Tyson rubric prior to the contest you judged? This could include receiving a copy of the rubric prior to the contest.

H: We received the copy of the rubric and a copy of what the score sheet would look like, and then at the site the director or the assistant director went over it with us.

I: Okay, are there any benefits of Tyson Rubric for show choir adjudication?

H: Oh, definitely. Because we all need to know. It helps with growth of the ensemble. It helps with development of the program as well. And it kind of gives you a gauge... well here we were two years ago, we were at this point and now here we are two years later - look at the growth that we had from this.

I: okay that's great, you're the first person that's mentioned that, so that's really cool. Did you detect any limitations of the Tyson Rubric?

H: No.

I: Okay, do you feel that the categories were appropriately weighted, because if you'll remember some certain aspects, received maybe 10 points and then others received five or 7? So, did you think that those were appropriately weighted and if not, what weights might be more effective in giving an appropriate balance of importance between categories?

H: From what I remember, I think it was appropriately weighted because I think the singing and the singing was the most important part, but then the theme and everything else... everything seemed even. It didn't seem like one mattered more than the other.

I: Did you have any other thoughts concerning the use of or the rubric itself?

H: No.

I: Are there any other issues concerning show choir scoring in general, that you think should be addressed in a rubric that maybe aren't right now?

H: No, not that I can think of because you talk about tone, and intonation, as well as the performance aspect and the overall theme.

I: Do you have any other thoughts, either about this specific rubric or show choir scoring in general, positively or negatively that you would like to comment on?

H: No. Positively, the thing that I like is that, sometimes in show choir we can hear the band overpower the choir. I think your rubric actually allows them to judge that aspect of the group of the show as well, because it actually matters. Because you can control that as a director before the show, you get a run through and sometimes you... some rubrics don't allow you to control that, to judge on that. I think that's a great thing.

I: Okay, I'm going to ask one more question. Several... I've had a lot of feedback regarding the types of judges that are judging the categories and there's some different opinions about this, and I would just like to have yours. So, there's a school of thought

that perhaps people who have vocal expertise should be judging the vocal portion of the sheet and people that have dance or visual expertise should be judging the visual expertise, and that should not overlap. And I was wondering, as an adjudicator how you feel about that issue?

H: I disagree, because as a... As a vocal teacher, as a singer and I've been in shows, I've been in choirs, I've been in the performance aspect of dance and I've taken dance classes as well, so I know how a show should run... And I know what it looks like if something looks off. I have enough knowledge as at a performer to know if something looks wrong. And just me judging on voice alone, that you're not utilizing your judge at the best of their abilities.

### **Appendix C.9: “Ms. Beech”**

(I = Interviewer; B = “Ms. Beech”)

I: Okay, can you hear me okay, Katie? Yeah, okay, I'm just going to start by reading the script. Thank you for taking the time to answer a few questions regarding your perceptions of the Tyson rubric. This interview is being recorded and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential, by continuing on, you are giving your consent to the interview and the publishing of the resulting data. Are you okay with all of that?

B: That's alright.

I: Then we're going to go ahead.

I: Could you tell me your current occupation? And you could also follow that up with your previous occupation.

B: Okay, currently, I am a full-time stay-at-home mom and a part-time freelance clinicians/adjudicator.

I: Okay.

B: Previously I was the director of Choral music at Jackson Academy in Jackson Mississippi for seven years.

I: Perfect could you briefly describe the types of score sheets that you have used at show choir contest? Are the rubrics global score sheets, weighted score sheets? And when I say global, I mean maybe you give one score, that's from zero to 100 and you circle the best number.

B: Okay, so in the past have used you tell me what I'm saying doesn't make any sense, but in the past, I've used I guess a global score sheet type of thing, where it would a category and then one through 10 and you would have to circle eight or nine or whatever. What have you? And they just put out to the side with no other explanation, and then I don't know that I've used a weighted? What is a weighted?

I: So, weighted would be the one that's currently in use in Mississippi, the one that you used this past year where some of the categories or sub-categories are worth more points than others.

B: Yes, okay, so I have in the creation of this particular rubric and it's to be a type of rubric, I have used a weighted rubric, but before that, I had not used either.

I: Okay, do you feel that for any of those score sheets that you have used the total points available for scoring how do you feel about that? Are they too wide to shallow? So, are there enough points, or not enough points to be adequate?

B: I think in most cases, they were plenty of points? I've never felt like there weren't enough to go around. Actually, the only time I have felt that way was pre-rubric, typical finals scoresheet. Sometimes you would only see 50 points at certain contests. Sometimes you might see 100, but I always felt like those were a little lacking but I like that better now that we've all adopted the rubric.

I: Do you feel that the evaluation sections meaning the vocal performance sections, the visual performance sections etcetera, need to be weighted or that they should all be worth the same amount.

B: I actually, like the weighted system? Because I do agree that there are certain things that are more important than others. And I like that it's just taking care of for us at the very get-go of the score sheet so we don't have to worry about adjusting on a 10- point scale.

I: Okay, the next question, could you speak to the problem of bias in choral adjudication?

B: How long you got? So, it's kind of how I feel about, like when I took an English class. You know when you take math, it's very clear either got the answer right or you didn't... You a teacher making to be similar ways, you're doing the work correctly, before the, the answer, but there's not a lot of wiggle room. But in English, you have how the teacher feels about what you wrote... Yes, there are some very clear things. I didn't put a comma here, and I should have, but most of the time it's very subjective, very whatever their opinion is, and I feel the same way about choral and show choir adjudication. You are paying for these people's opinions but a lot of times, unfortunately, you also get their bias. I feel it's very hard to find judges consistently, let's put it that way, who don't have feelings One way or another towards a director or a certain group or judges who could turn that off, because we all feel it, but we have to be able to turn it off, right?

I: Do you feel that the type or does the type of score sheet, used to have any impact on adjudicator bias?

B: Absolutely, like I, I think you can... I think that's one of the only ways to combat it is. Well, you have the first way, which is try to best to hire people who you don't feel like are going to be biased, right? And I think if you just give them a loosey-goosey score sheet, where they can just put somebody where they want to and all the, they don't owe any explanation. You know, it's just asking for trouble. But if you give them this kind of rubric style, they kind of have to answer to that. It's very clear what their score should mean.

I: Okay, could you speak to the usability of the Tyson rubric from an adjudicators point of view?

B: Yes, I adore it. I think it's the best. I like that it's on the iPad or on the computer and that you can click through and doing, you do. So, I like that part, the technological part... But I also think it's good because once you... It may take you a few minutes beforehand to really study the rubric and know what you want... What's expected? But after that, it's very simple, and it keeps it really streamlined, I think group-to-group, but also gives you enough wiggle room where you don't have to slaughter a group to get your point across and say "Hey I want you to be a five, but here You are, you're a three. And what I like is when that Director sees the here, say in the three, there's an explanation. That's my favorite part. But in terms of ease of-use, it's very easy, to me.

I: How does the Tyson rubric compare to other score sheets, you have used?

B: It's just more in-depth, it's more in-depth in terms of explanation of set scores. And I like, as I've said previously, I like the weighted aspect, so that it's clear that certain things are obviously more important than other things and it doesn't give, those less important categories the same weight to that group's overall score.

I: How does the Tyson Rubric compare to other score sheets that you have used, maybe that aren't weighted, rubrics?

B: Compare and... Just in generally?

I: Yeah.

B: As I previously mentioned, the only thing that people could maybe say, "Oh this isn't so great about it. Is that take a little study time beforehand in my opinion, you should... So, you should take that rubric for two or three days and then you really look over it, so you're well-versed before you open up the program. So, in terms of that yeah, it takes a little more time, little more involved on the front end, but then it's just clear there's just clarity for the judge. There's clarity for the director that's receiving the scores and there's clarity for the host of the competition, which I also think is important, because a lot of times that's the person that gets the junk, right, they're the ones that get the complaints, not really the judge, and so it gives them just set a clear conscious, about what they're presenting to both sides.

I: Okay, what orientation or training did you receive for the Tyson rubric prior to the contest that you judged? This could include receiving a copy prior to the contest.

B: Okay, yes, almost every time if not every time I've been asked to use that rubric I've received a copy ahead of time to study and then I've also been a part of a hands on training I guess in the program before we ever go into the auditorium.

I: Okay, great, are there any benefits of the Tyson rubric for show choir adjudication?

B: As previously mentioned clarity for all parties to where you know what your score means and you know what you're communicating to that group. other benefits as a judge, it kind of releases me to really be able to express what I need to express to those groups beforehand. If I was judging and you don't have any kind of guidance to what a 5 is supposed to mean and what a 7 is supposed to mean, you would either score got maybe too high for fear of hurting their feelings or other judges would just bottom them out and give them ones and twos. And then that poor group, just thinks that they're just the pits, right? And so, I think that one of the biggest benefits is that as a judge, I can go in with no weight, on my shoulders, I can see the group for what it is and then take my opinion and my education toward it and find what it says on the rubric and give them that score.

I: Did you detect any limitations of the Tyson rubric?

B: Not so far, I don't recall any.

I: okay, did you think that the categories of the Tyson Rubric were appropriately weighted? And if not, what weights might be more effective in giving appropriate balance of importance?

B: Each time I've used it, I've felt that they've been appropriately waited. I don't recall anything needing changing.

I: Do you have any other thoughts of maybe questions that I didn't ask about the rubric, itself?

B: So, I think it's very clear, very well-presented, and it's very helpful.

I: Do you think that there are other issues concerning show choir scoring that should be addressed within a rubric?

B: Yeah, is there a spot anywhere there, or Is it just per contest, about Time limitations and certain things like that?

I: it's outside of the scope of the rubric. I think that there are those rules but that it's not in the rubric itself, but you could certainly... My next question, it's just other thoughts that you have about show choir scoring in general, positive, negative other things that need to be addressed.

B: I think that since we've gone with unification in so many things, it might not be a bad idea to extend that to certain contest rules that we all agree should be in place.

You know every contest's going to have a few different things, but things like a time limit if everybody across the state or South East or whatever said, Okay, it's going to be 20 minutes, 25 minutes on and off. I think that that could then be translated into the score sheet. So, it's not flippable. Yeah, because I feel like we all put that in package, but then how strictly is it enforced? And then you have groups that are coming in and taking 30 minutes on stage, and at some contests, they're going to be docked for it and in other contests they just turn a blind eye.

Continuing to fight for complete fairness across the board in terms of the only difference that should be in your score or the only thing that should set you apart at the end of the day is... Did you sing better, or did you dance better, did you perform better? So, then it's not so much a thing. Well, we have these huge sets that is going to take 35 minutes but we didn't get any penalties...

I: Right, so maybe adding a place on the actual score sheet and in Carman where those penalties are deducted not by somebody backstage but perhaps by a judge?

B: Correct, or are done on that side of things, where it's a little more official right?

I: That's a really good idea. Nobody has said that yet. I like that.

### **Appendix C.10: “Mr. Byrd”**

(I = Interviewer; BY = “Mr. Byrd”)

I: Perfect, okay, thanks for taking the time to answer a few questions regarding your perceptions of the Tyson rubric. This interview is being recorded and will then be transcribed and coded. You will not be identified in the resulting data and all of your information will be confidential, by continuing on your giving your consent to the interview and the publishing of the resulting data is that okay with you?

BY: Absolutely:

I: okay, this should only take about 10 minutes. So, I'm just going to go through the questions, first of all, could you tell me your current occupation?

BY: I am the Assistant to the Dean for Arts PR and Marketing at Southern Miss and I'm the artistic director of Festival South.

I: Could you also speak just a little bit about your experience in the band marching world and your adjudication experience with that.

BY: Sure, I was a band director for a decade in Texas, I taught in Mississippi prior to that, just as I was doing schooling. At some point I became an adjudicator for the Texas color guard circuit, which was the only color guard circuit in Texas at that time and eventually that led to Winter Guard International, judging for them, which led to Drum Corp International, which led to Bands of America. So, I was judging for all the major sort of marching arts, I guess, you could call it Organizations. I was Chief Judge in Texas for about 10 years, not quite because it was not long after I got there, and then I stayed that way even though I was here in Mississippi. They liked somebody was sort of out of state to deal with any technical issues in all and not connected to anyone there. So, I did actually for a while, even after I left the state and I guess that's mostly it. I still continue to adjudicate Winter Guard, all over the world actually. And I do some marching band, I just did some marching band this year. It just depends on my schedule works with that. I guess the most recent thing that happened to me as I was inducted in the Texas color guard Hall of Fame.

I: Oh congrats, I didn't know that.

BY: Yeah, yeah, no, it's kind of shocking to me.

I: but that's awesome.

BY: I feel like I need to be 80 for that to happen, but I don't know.

I: Who knew that you could accomplish that at 29?

BY: Yeah, exactly. I have been judging things of all types, and I was actually the education person as chief judge, so I trained judges, all through, Texas for years.

I: Great, well with that experience, could you briefly describe the types of score sheets that you've used, specifically at show choir contests? Are they rubrics if so are they weighted or are they... Or have you used global score sheets, meaning here's a score from one to whatever this number is and you just circle a number? If you could just speak a little bit about that.

BY: So I have very little experience with show choir, but the few that I have done, one was years ago, and it was so wide open that I recall it just being like I was like, "I don't know how to judge this after judging for years that it was a scale for each area so it was like a one-to-10 kind of thing, for each area and there were 20 areas or something that

you were supposed to catch during the performance. The last one I did, I'm not sure if it was yours or not, that we used that was sort of a rubric sheet that did describe qualifications or each scoring zone. So, a one to three was this, a four to six was that, that sort of give you a little bit better idea of where to be? However, there were still tons of categories with each rubric-based item, and every judge had the same thing so everybody was sort of going after the same thing, which is sort of different than anything I have been used to in the other arts. It was a simpler idea, but just a way broader...

I: We may come back to that at the last question, because I think you have a unique perspective. Do you feel that evaluation sections should be weighted meaning should vocal category received the same amount of points as visual category or should one be worth more than the other or should they all receive the same amount of points?

BY: I strongly believe that this is a music activity, and the teaching and training of music should be the priority. Now whether that's 60/40, 70/30, 80/20, there's that. I know that for years, the Texas UIL which is like their State Marching everything had a total of five judges three were music, two were visual and actually they were just marching judges, so it was technical marching. It almost didn't matter what your show was, as long as you marched it cleanly. So it did skew music. Okay in Winterguard world and of bands of America world. And let's do Drum Corp and Bands of America. The wind side is skewed a little bit, so the actual music side has a single more judge than the visual side, so it is pushed that way, say, in Winterguard world since it's mostly a visual thing, it's about the technique. So, there are two judges who are really about technique, movement and equipment. So, the core of whatever the activity is, I think, should be the focus.

I: Okay, great, could you talk about the problem in your experience of bias in choral adjudication?

BY: Oh wow...

I: if you perceive that there is one...

BY: absolutely, I think several things happen. I think people have different ideas of what a sound should be. I don't think there's a qualification or anything for size of group and how that affects the sound because I've heard some amazingly accurate wonderful tiny groups that just can't seem to compete for with something that you know 100 voice, 50 voice choir kind of thing, that is actually singing much sloppier, but there's just an impact to it. So, I think I would say the... Sometimes our eyes also mislead us. The better dress shows your production, you just get wowed in the moment and when you go back and listen to it, you're just like, Wait, that wasn't great. So, I don't know if that's bias or just as sort of showmanship overwhelms the actual technical aspect of singing. I do know that it has been shocking to me, who's allowed to judge show choir things sometimes. In Texas, man, it is such a system to even become a Judge, it's two years of training to do that kind of thing. And, you, any affiliation that you have with anybody that you have worked with or done, and you can't judge. You're just not on that panel. And so, it's a hard thing to deal with obviously because suddenly you're bringing in judges from a long way away, and it gets expensive. And I realize that side of it too, but I think there is some... I've kind of myself when I know the people I'm actually harder on them, that's my bias, I think, and I have to be really careful about that, because I feel like they know better. How could they do this? Whereas I would give other groups, the benefit of the doubt, really quickly. So I guess that's a different kind of bias as well.



I: Okay, does the type of score sheet used have any impact on adjudicator bias?

BY: I believe so, I think the wording of especially if you're using a rubric, the wording of that rubric if people have not been trained to understand what those words truly mean and that happens in all our activities all the time is that people interpret we like to use in guard world, "visual musicality", which means, do I see what I hear? And people interpret that all sorts of ways, many times, and they just read into it way more than wait, that's not what I'm hearing. So I think without some level of training, there could be some bias as to what the words actually mean on a rubric kind of sheet, and I'm sure there are other ways that just the way things are worded, I will tell you that I was in a process that it's ongoing where those sheets were updated every year or sometimes during a year of somebody would say, "Well that doesn't mean what you think it means" and say, we'd adjust it so that it became a little bit more global, but I'm sure there are ways like that, but I think it may come from more either how it's written or the understanding of what's written.

I: Okay, could you speak to the usability of the Tyson Rubric from an adjudicators point of view?

BY: I thought it was spot on and I think it's what I'm most familiar with where it clarifies scoring ranges, it clarifies what needs to be done. The only, I think, exception that I had with it was it's too broad for a show choir thing and my recommendation would be that there'd be somebody that focuses just on movement, then somebody that focuses on the general effect of everything, which gets into the costuming, the sets, the flow of the production, the design elements of that and then maybe the others were specific to music. I just felt like it was overwhelming the amount of things you had to sort of take in. And in my world, you as somebody who taught judges for a long time, you have to comment on all of those things on your tape or you're really not allowed to score it. So, in my thinking when I found that sheet first time I was like, Holy Smoke, I've got to say something about all of these things, and I found myself having to look away a lot to see what it was I needed to do next because there were so many areas of adjudication. So if I had to say anything, I think it's on the right track, but I think it just needs to be maybe a little bit more... And that's the activity as a whole. That may not be the rubric. It's overwhelming amount of information for one person to really, I think adjudicate effectively.

I: Okay, that's great, I think in that answer you actually answered the next two questions. Could you say, maybe how the Tyson rubric compares to other score sheets that you have used and you've kind of talked about that a little bit, but if there are something else that you wanted to add.

BY: No, I think it seems right on track with other things. Like I said, I think just something... I think it would be wonderful to test in this activity. A contest where, like I said, there were a movement judge and they were an effect judge and then maybe there were three music judges and you might even one of those music judges be something that's even more specific that's not just music, music, music, but it's music training and quality, and that not just the selection and how it works and all those things that maybe it's actually a technical singing kind of judge. That's a hard category, though. So, the other thing would be is just a look at some other sheets and other activities and how they're done. I feel like Winterguard, who started amazingly enough, all of the judging

and book and sheets and all of those things, and then DCI and BOA sort of adopted them all because they realized they were using all of the same judges. But maybe look at the formatting of those sheets, because I know what they do, is they put all of the descriptors on the back of a sheet, the full out descriptors and then on the front of the sheet are sort of the meatiest words per caption. So, gosh, I don't have something in front me just to say it, but the whole paragraph description of what is in as we would call it box three, which is scoring from whatever to whatever, is on the back of the sheet but on the front of the sheet it just lists overall what you're judging because it's interesting enough, in the sheets, like for Winterguard and all is the words are very much the same of what you're looking for in every box. The adjectives of how it happens. So, in like, they divide stuff into what you're doing and how you're doing it. And on the front side they say zero is infrequently achieves whatever. And then in the next box up, it says sometimes, in the middle box is usually for boxes always, and then the top is top in scoring, which is box five for us, it says exceeds those things. So, the descriptors are the same, then it could be the tone, the dancing, the arrangements, the costume, the this, but it could be doesn't. So those keywords the tone, the dancing, are on the front of the sheet where you're actually scoring and looking at, but then the descriptors to fully enhance that. And I know when I'm judging I only use the front of the sheet and I'm trying to talk about all of those things, and then I turn it over. And you do a sort of analysis of where they always doing this, were they sometimes, were they exceeding all of these things and it sometimes a variation, so it does a score. So, we used to call it in judging world "impression analysis comparison" for the three steps of judges judging. So, when you're watching you get an impression of where they fall on that scale and then you go to the sheets and you do an analysis to refine what that is and before you put that final score down, is you have to make a comparison. If I put this number down, does that make them way better than their competition or slightly better or equal to or less than... So, it's those three steps of judging that you sort of go through for every group. But I think having the sheet and that kind of organization might be helpful. And I'm trying to remember what your sheet was, it seems like there was a lot of information at once.

I: What orientation or training did you receive for the Tyson rubric prior to the contest you judged? This could include receiving a copy of the rubric prior to the contest.

BY: I got a copy prior to the contest. I think you sent, so it's been, like, a year now, you sent a follow up about like, "Hey, look at this and check this". And then prior to the contest, the contest director who I think we're all friends with went through and said, "Hey take a look at this. Again, this is a little different, we're doing this: , because there were several seasoned judges on the panel that had not used this before but they had had a chance to look at it, and I think the general feeling of it was... Wow, this is way better. Especially those that had used other things before.

I: And then I'm going to skip a couple because you've already really talked about those. As far as the categories go for the Tyson rubric, if you remember, do you feel that they were appropriately weighted, and just as a refresher vocal was probably about 60% of the score, visual 30, show 10... I think, is how that percentage-wise worked out, and also some of the sub-categories were worth more than others. For example, vocal tone was worth more than style. And then in the visual category execution was worth more than

costumes. So, do you feel that that's appropriately weighted or is there a better way to do that?

BY: I remember thinking that it seemed to make sense in the moment that it seemed that way and that it was skewed towards music. Like I said, the only thing I really recall was that it was just an overwhelming amount of things to evaluate at once.

I: Is there any other thoughts you have about the rubric itself that you haven't shared that you would like to share?

BY: No, I think I've covered all.

I: Okay, how about any other issues about show choir scoring that you feel that should be addressed in a rubric that is not currently addressed in a rubric?

BY: Other than what I said that I think the evaluation of the groups would be so much better, and at the end of the day, so much more helpful to the students and the teachers if a movement professional were judging movement and nobody else was really in on that. If a design professional that does theatrical productions or shows of different things were doing a design show element, and that musicians focused on music, what they're doing, I think they'd get a lot better feedback. And at the end of the day, show choir works a lot like Winterguard is that it's not this one-off, and the drum corps, all of those things. It's not a one-off competition where... Congratulations, you won, and that's it, and it's done. It's sort of a process that you go, ooh, they suggested this, maybe you could try this and we, we make it better, which is a great lesson, I think, for students too, is that it's always a growth process the things you can get better and that lots of feedback is valuable even if you don't agree with it sometimes, it's valuable. I think the reason I was successful with a lot of what I did was we trained people to talk to the students and let it be the extra teacher in the room, that so many programs don't have. And you know that if you go to a contest, and suddenly you listen to those judges tapes and you've been saying that over and over and over and over or have a guest clinician in and everything you have been saying, suddenly they act as if it's gold when somebody else says it. And so being able to have that possibility, I think is a really strong thing, but when they also know and trust the system a little bit more, knowing, wow - this is a professional dance person telling us this was good, or you could fix this by doing whatever. And so that even though I've had some days training, and worked in that world a little bit, I'm not the person to tell them, ooh your balance on this and your point of the da, da, da, not am I the person to say your vocal projection is impaired because you are doing a blah blah and whatever. But I am probably the person to tell them the costumes work like this; the show works like this. You set out your theme to be this. And so, I think if I were able to really play to my strengths, I could be more helpful. So that would be the... And I realized that this system wide major changes.

I: It is but there are places that do that, some of the not local, but regional or national competitions, do that, and I've actually judged that before as the vocal expertise where I didn't look at all or score at all on visuals or overall which was different for me, but I liked it. So, it would have to be one of those things where we either we all do it in Mississippi or nobody does it. But I think that we're... I think this is helpful, because I think you're the third person out of the four interviews I've done so far, that has had that same feedback.

BY: What I mentioned about UIL Texas, that it was three judges, three music and to marching, so I understand that this year, because Texas is slow to change, especially in very esteemed marching band world, they have changed it and there are three music judges, but now there are two basically music effect judges, who judge marching but also program design and effect, and this. And I'll be honest with you, years ago, and I mean this will tell you how long ago this was... so I was still living in Texas or maybe it was the first year I moved. So we're talking about 2002. I was on a test panel at state champs marching finals for Texas which is like, "Oh my God, that's hallowed ground, like Holy crap. But they had their regular three and two panel, but they did a five-person panel like bands of American or drum corps do for judging, to see what the results would be. No one ever saw those results, because I don't think they wanted to know because some bands that placed really well were bands that were doing very little, but doing it cleanly. And so, a band that was some of the really progressive shows that people are doing now and they're dancing and doing all these different things were not getting scored the same way in that system of it's clean or it's done. And that was it. And there was in a year that I had a band in state finals, a band that got first across the board in music and 20th across the board in marching and still won. Because the way they do it is they total ordinals there, so like what the judges do a sheet that's another thing, I don't know, if it gets into your ears about the rubric, but how the rubric is determined. So, do you do total points and the points win? So, in Texas for those, it's the ordinals win. Which means if there five judges... And you were first, first, first, fifth, tenth, that means you're actually 18, your ordinal is 18. So, 1,1,1,5,10, and then... And so, if you do that to everybody the ordinal wins which I think is way more fair because imagine if somebody just especially depending on the range of numbers you have in it, a contest, that's larger, the larger they get, the tighter they get, and suddenly one judge can throw the whole contest. I think you're 20 points better than them in this category. And so, that 20 points in a tight contest could cost somebody first place or 20th place. So, the ordinal system I think is a really smart way to make that happen.

## APPENDIX D – ISSMA Rubric

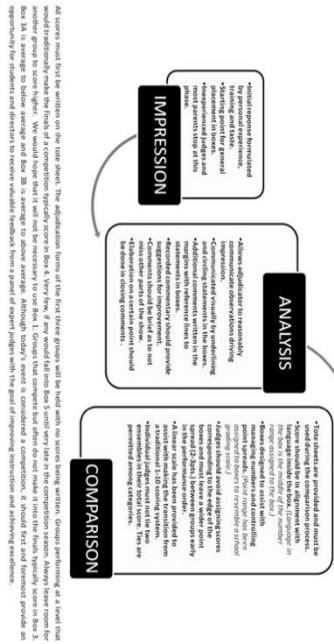
# ISSMA SHOW CHOIR QUALIFICATION



## Instrumental Performance

### INDIANA SHOW CHOIR QUALIFICATION CRITERION REFERENCE JUDGING FORMAT

The Indiana State School Music Association has developed a criterion reference judging system that includes dimensions from a musical grouping, style and traditional show choir instrumentation for all sanctioned show choir institutions. This system requires each judge to be familiar with the language in each of the boxes and the corresponding point values. The language in boxes should be used by the judges when making a recorded analysis of each group's performance. Each judge is assigned to evaluate one of three categories: VOCAL, VISUAL, or INSTRUMENTAL. Each section uses a limited rubric compared to three categories with a scoring range of 1-100 making a total possible score of 300 points. The three categories are Technique, Artistry, and Repertoire Effect. Repertoire Effect has been added to allow judges to better evaluate the stage elements and effectiveness of each performance.



## ISSMA SHOW CHOIR QUALIFICATION INSTRUMENTAL PERFORMANCE

	TECHNIQUE (100 pts.)					ARTISTRY (100 pts.)					REPERTOIRE EFFECT (100 pts.)									
	Box 1 (0-30)	Box 2 (31-60)	Box 3A (61-70)	Box 3B (71-80)	Box 4 (81-95)	Box 5 (96-100)	Box 1 (0-30)	Box 2 (31-60)	Box 3A (61-70)	Box 3B (71-80)	Box 4 (81-95)	Box 5 (96-100)	Box 1 (0-30)	Box 2 (31-60)	Box 3A (61-70)	Box 3B (71-80)	Box 4 (81-95)	Box 5 (96-100)		
	TECHNIQUE: The instrumental ensemble exhibits <b>very little</b> understanding of good tone quality or articulation. *Instrumental parts are <b>not balanced</b> . *Instrumental parts are <b>not in tune</b> . *Pitch inconsistencies occur and are <b>not</b> corrected. *A majority of the ensemble displays <b>no</b> skill with regard to rhythmic and/or note accuracy. *Ensemble cohesiveness is <b>not</b> achieved.	TECHNIQUE: The instrumental ensemble exhibits a <b>limited</b> understanding of good tone quality and articulation. *Good balance is <b>seldom</b> achieved. *Instrumental parts are <b>seldom</b> in tune. *Pitch inconsistencies occur and are <b>rarely</b> corrected. *Individuals within the ensemble display <b>significant</b> lack of skill with regard to rhythmic and/or note accuracy. *Ensemble cohesiveness is <b>rarely</b> achieved.	TECHNIQUE: The instrumental ensemble exhibits <b>an avg. to below avg.</b> approach to tone quality and articulation. *Balance is <b>an avg. to below avg.</b> within the ensemble with frequent inconsistencies being apparent. *Instrumental parts are <b>not</b> always in tune. In general, intonation is <b>an avg. to below avg.</b> *Pitch inconsistencies occur and are <b>not</b> corrected. *Individuals within the ensemble display <b>an avg. to below avg.</b> skill with regard to rhythmic and/or note accuracy with cohesiveness being <b>an avg. to below avg.</b>	TECHNIQUE: The instrumental ensemble exhibits <b>an avg. to above avg.</b> approach to tone quality and articulation. *Balance is <b>an avg. to above avg.</b> within the ensemble but inconsistencies are apparent. *Instrumental parts are <b>not</b> always in tune. In general, intonation is <b>an avg. to above avg.</b> *Pitch inconsistencies occur, but are <b>often</b> corrected. *Individuals within the ensemble display <b>an avg. to above avg.</b> skill with regard to rhythmic and/or note accuracy. *Ensemble cohesiveness may <b>not</b> always be sustained and is <b>an avg. to above avg.</b>	TECHNIQUE: The instrumental ensemble exhibits <b>a very good</b> control of tone quality and articulation. *Very good balance within the ensemble is maintained throughout most of the production. *Instrumental parts are <b>usually</b> in tune. *Pitch inconsistencies occur, but are <b>quickly</b> corrected. *Ensemble displays <b>very good</b> skill with regard to rhythmic and/or note accuracy with <b>very minor</b> breakdowns in cohesiveness.	TECHNIQUE: The instrumental ensemble exhibits <b>superior</b> control of tone quality and articulation. *Outstanding balance within the ensemble is maintained throughout the entire production. *Instrumental parts are <b>always</b> in tune. *Pitch inconsistencies occur. *Ensemble displays <b>complete</b> control of skill with regard to rhythmic and/or note accuracy with no breakdowns in cohesiveness.	ARTISTRY: The instrumental ensemble exhibits <b>very little</b> understanding of the artistry of music performance. *Dynamics and expressive shaping of phrases are <b>not</b> used to enhance the music. *The ensemble <b>does not</b> reflect an understanding of style or idiomatic interpretation. *Performers <b>do not</b> demonstrate any understanding of professionalism with frequent major lapses occurring in concentration, discipline, and engagement.	ARTISTRY: The instrumental ensemble exhibits a <b>limited</b> understanding of the artistry of music performance. *Dynamics and expressive shaping of phrases are <b>seldom</b> used to enhance the music. *The ensemble <b>seldom</b> reflects an understanding of style or idiomatic interpretation. *Performers <b>seldom</b> demonstrate an understanding of professionalism with major lapses occurring in concentration, discipline, and engagement.	ARTISTRY: The instrumental ensemble exhibits <b>an avg. to below avg.</b> achievement level in the artistry of music performance. *There is <b>an avg. to below avg.</b> use of dynamics and expressive shaping of phrases to enhance the emotional aspects of the music. *The ensemble reflects <b>an avg. to below avg.</b> approach to style and idiomatic interpretation. *Performers demonstrate <b>an avg. to below avg.</b> understanding of professionalism with frequent lapses occurring in concentration, discipline, and engagement.	ARTISTRY: The instrumental ensemble exhibits <b>an avg. to above avg.</b> achievement level in the artistry of music performance. *There is <b>an avg. to above avg.</b> use of dynamics and expressive shaping of phrases to enhance the emotional aspects of the music. *The ensemble reflects <b>an avg. to above avg.</b> approach to style and idiomatic interpretation. *Performers demonstrate <b>an avg. to above avg.</b> understanding of professionalism, but lapses occur in concentration, discipline, and engagement.	ARTISTRY: The instrumental ensemble exhibits <b>a high level</b> of achievement in the artistry of music performance. *Dynamics and expressive shaping of phrases are <b>usually</b> used to enhance the emotional aspects of the music. *The ensemble <b>usually</b> reflects a mature and superior approach to style and idiomatic interpretation. *A <b>high</b> degree of professionalism is present, but <b>minor</b> lapses in concentration, discipline, and engagement occur.	ARTISTRY: The instrumental ensemble exhibits <b>superior</b> achievement in the artistry of music performance. *Dynamics and expressive shaping of phrases are <b>always</b> used to effectively enhance the music. *The ensemble <b>always</b> reflects a mature and superior approach to style and idiomatic interpretation. *The <b>highest</b> degree of professionalism is <b>constantly</b> present.	ARTISTRY: The instrumental ensemble exhibits <b>superior</b> achievement in the artistry of music performance. *Dynamics and expressive shaping of phrases are <b>always</b> used to effectively enhance the music. *The ensemble <b>always</b> reflects a mature and superior approach to style and idiomatic interpretation. *The <b>highest</b> degree of professionalism is <b>constantly</b> present.		REPERTOIRE EFFECT: Repertoire provides <b>very little</b> opportunity to enhance the effectiveness of the instrumental performance. *The audience is <b>not</b> engaged due to the lack of creativity, transitions, and variety of the instrumental elements of the production. *Construction of the repertoire is <b>extremely out of balance</b> with the demonstrated level of performance. *Proper balance is <b>not</b> achieved between the vocal ensemble and the accompaniment. *The inter-relationship between the instrumental and the vocal and visual elements is <b>not</b> effective.	REPERTOIRE EFFECT: Repertoire provides a <b>limited</b> opportunity to enhance the effectiveness of the instrumental performance. *The level of involvement of the audience is <b>limited</b> by the lack of creativity, transitions, and variety of the instrumental elements of the production. *Construction of the repertoire is <b>seldom</b> well balanced with the demonstrated level of performance. *Proper balance is <b>rarely</b> achieved between the vocal ensemble and the accompaniment. *The inter-relationship between the instrumental and the vocal and visual elements provides a <b>limited</b> level of effectiveness.	REPERTOIRE EFFECT: Repertoire provides <b>an avg. to below avg.</b> opportunity to enhance the effectiveness of the instrumental performance. *The audience is <b>occasionally</b> entertained with the level of involvement being influenced by <b>an avg. to below avg.</b> degree of creativity, effective transitions, and variety of the instrumental elements of the production. *Balance between the construction of the repertoire and the demonstrated level of performance is <b>an avg. to below avg.</b> *Proper balance is <b>low</b> but <b>moderately</b> maintained with the vocal ensemble. *The inter-relationship between the instrumental and the vocal and visual elements provides <b>an avg. to below avg.</b> level of effectiveness.	REPERTOIRE EFFECT: Repertoire provides <b>an avg. to above avg.</b> opportunity to enhance the effectiveness of the instrumental performance. *The audience is <b>often</b> entertained with the level of involvement being influenced by <b>an avg. to above avg.</b> degree of creativity, effective transitions, and variety of the instrumental elements of the production. *Balance between the construction of the repertoire and the demonstrated level of performance is <b>an avg. to above avg.</b> *Proper balance is <b>moderately</b> maintained with the vocal ensemble. *The inter-relationship between the instrumental and the vocal and visual elements provides <b>an avg. to above avg.</b> level of effectiveness.	REPERTOIRE EFFECT: Repertoire provides <b>an excellent</b> opportunity to enhance the effectiveness of the instrumental performance. *The audience is <b>constantly</b> entertained and kept engaged by the creativity, effective transitions, variety and pacing of the instrumental elements of the production. *Balance between the construction of the repertoire and the demonstrated level of performance is <b>excellent</b> . *Proper balance is <b>always</b> maintained with the vocal ensemble. *The inter-relationship between the instrumental and the vocal and visual elements provides <b>maximum</b> levels of effectiveness.	REPERTOIRE EFFECT: Repertoire provides <b>an exceptional</b> opportunity to enhance the effectiveness of the instrumental performance. *The audience is <b>constantly</b> entertained and kept engaged by the creativity, effective transitions, variety and pacing of the instrumental elements of the production. *Balance between the construction of the repertoire and the demonstrated level of performance is <b>excellent</b> . *Proper balance is <b>always</b> maintained with the vocal ensemble. *The inter-relationship between the instrumental and the vocal and visual elements provides <b>maximum</b> levels of effectiveness.
	1 - 2	3 - 4	5	6	7	8 - 9	10	TOTAL SCORE (300 pts possible)												
	Comments																			
	TOTAL SCORE (Possible 300 pts)																			
	More than one adult – Do not post on tabulation form.																			

More than one adult – Do not post on tabulation form.

Judges Signature \_\_\_\_\_



## ISSMA SHOW CHOIR QUALIFICATION VISUAL PERFORMANCE

	Box 1 (0-30)	Box 2 (31-60)	Box 3A (61-70)	Box 3B (71-80)	Box 4 (81-90)	Box 5 (90-100)	TECHNIQUE (100 pts.)
TECHNIQUE	TECHNIQUE-Performers exhibit <i>very little</i> understanding of the technical aspects of the visual program. * Alignment and spacing is <i>rarely</i> good. * The body carriage, body control, and posture reflect a <i>limited</i> degree of muscle control. * Props and/or equipment are handled with <i>no precision and control</i> . * Performers demonstrate <i>no skill</i> in regard to dance fundamentals.	TECHNIQUE-Performers exhibit a <i>limited</i> understanding of the technical aspects of the visual program. * Alignment and spacing is <i>fair</i> and flaws are <i>frequent</i> . * The body carriage, body control, and posture reflect a <i>developing</i> degree of muscle control, but are not consistent throughout the ensemble. * Props and/or equipment are handled with <i>only a fair</i> degree of precision and control. * Movement uniformity is <i>not consistent</i> . * Performers demonstrate <i>significant</i> lack of skill in regard to dance fundamentals.	TECHNIQUE-Performers exhibit <i>avg. to below avg.</i> control of the technical aspects of the visual program. * Alignment and spacing is <i>avg. to below avg.</i> , with flaws being evident and very slow recovery. * The body carriage, body control, and posture reflect an <i>adequate</i> degree of muscle control, but may not be consistent throughout the ensemble. * Props and/or equipment are handled with a <i>less than moderate</i> degree of precision and control. * Movement uniformity is <i>avg. to below avg.</i> . * Performers demonstrate <i>avg. to below avg.</i> skill in regard to dance fundamentals.	TECHNIQUE-Performers exhibit <i>avg. to above avg.</i> control of the technical aspects of the visual program. * Alignment and spacing is <i>avg. to above avg.</i> , with flaws being evident and slow recovery. * The body carriage, body control, and posture reflect an <i>adequate</i> degree of muscle control, but may not be consistent throughout the ensemble. * Props and/or equipment are handled with an <i>adequate</i> degree of precision and control. * Movement uniformity is <i>avg. to above avg.</i> . * Performers demonstrate <i>avg. to above avg.</i> skill in regard to dance fundamentals.	TECHNIQUE-Performers exhibit <i>very good</i> control of the technical aspects of the visual program. * Alignment and spacing is <i>generally</i> good, with minor flaws, but with quick recovery. * The body carriage, body control, and posture reflect a <i>high degree</i> of muscle control. * Props and/or equipment are handled with <i>excellent</i> precision and control. * Movement uniformity reflects <i>maximum</i> precision and is maintained throughout. * Performers <i>always</i> demonstrate <i>complete</i> control of skill in regard to dance fundamentals.	TECHNIQUE-Performers exhibit <i>superior</i> control of the technical aspects of the visual performance. * <i>Excellent</i> alignment and spacing is maintained throughout. * The body carriage, body control, and posture reflect <i>superior</i> muscle control. * Props and/or equipment are handled with <i>excellent</i> precision and control. * Movement uniformity reflects <i>maximum</i> precision and is maintained throughout. * Performers <i>always</i> demonstrate <i>complete</i> control of skill in regard to dance fundamentals.	
ARTISTRY	Box 1 (0-30) ARTISTRY-Performers exhibit <i>very little</i> understanding of the artistry of the visual performance. * Performers are <i>not able</i> to communicate and portray emotion. * The visual performance <i>does not</i> reflect an understanding of style and idiomatic interpretation. * Performers <i>do not</i> demonstrate an understanding of professionalism with <i>frequent</i> major lapses occurring in concentration, discipline, and engagement.	Box 2 (31-60) ARTISTRY-Performers exhibit a <i>limited</i> understanding of the artistry of the visual performance. * Performers are <i>seldom</i> able to communicate and portray emotion. * Performers <i>seldom</i> reflect an understanding of style and idiomatic interpretation. * Performers <i>seldom</i> demonstrate an understanding of professionalism with <i>major</i> lapses occurring in concentration, discipline, and engagement.	Box 3A (61-70) ARTISTRY-Performers exhibit an <i>avg. to below avg.</i> level of achievement in the artistry of the visual performance. * Performers demonstrate an <i>avg. to below avg.</i> degree of communication and emotion through the use of facial expression and characterization. * The visual performance reflects an <i>avg. to below avg.</i> approach to style and idiomatic interpretation. * Performers demonstrate an <i>avg. to below avg.</i> understanding of professionalism with <i>frequent</i> lapses occurring in concentration, discipline, and engagement.	Box 3B (71-80) ARTISTRY-Performers exhibit an <i>avg. to above avg.</i> level of achievement in the artistry of the visual performance. * Performers demonstrate an <i>avg. to above avg.</i> degree of communication and emotion through the use of facial expression and characterization. * The visual performance reflects an <i>avg. to above avg.</i> approach to style and idiomatic interpretation. * Performers demonstrate an <i>avg. to above avg.</i> understanding of professionalism, but lapses occur in concentration, discipline, and engagement.	Box 4 (81-90) ARTISTRY-Performers exhibit a <i>high level</i> of achievement in the artistry of the visual performance. * Performers <i>usually</i> demonstrate a <i>high</i> degree of communication and emotion through the use of facial expression and characterization. * The visual performance <i>usually</i> reflects a <i>minor</i> and <i>very good</i> approach to style and idiomatic interpretation. * A <i>high degree</i> of professionalism is present, but <i>minor</i> lapses in concentration, discipline, and engagement occur.	Box 5 (90-100) ARTISTRY-Performers exhibit <i>superior</i> achievement in the artistry of the visual performance. * Performers demonstrate <i>outstanding</i> communication and emotion through the use of facial expression and characterization. * The visual performance <i>always</i> reflects a <i>major</i> and <i>superior</i> approach to style and idiomatic interpretation. * The <i>highest</i> degree of professionalism is consistently present.	ARTISTRY (100 pts.)
REPertoire EFFECT	Box 1 (0-30) REPertoire EFFECT-Repertoire provides <i>very little</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>not engaged</i> due to the lack of creativity, transitions, variety and pacing of the visual elements of the production. * Construction of the repertoire is <i>extremely</i> out of balance with the demonstrated level of performance. * Costuming and staging shows <i>little</i> creativity and is <i>not</i> effective. * The inter-relationship between the vocal and musical elements is <i>not</i> effective.	Box 2 (31-60) REPertoire EFFECT-Repertoire provides a <i>limited</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>seldom</i> entertained with the level of involvement being influenced by an <i>avg. to below avg.</i> degree of creativity, effective transitions, variety and pacing of the visual elements of the production. * Construction of the repertoire is <i>seldom</i> well balanced with the demonstrated level of performance. * Costuming and staging is <i>only somewhat</i> creative and is <i>seldom</i> effective. * The inter-relationship between the vocal and musical elements provides a <i>limited</i> level of effectiveness.	Box 3A (61-70) REPertoire EFFECT-Repertoire provides an <i>avg. to below avg.</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>seldom</i> entertained with the level of involvement being influenced by an <i>avg. to below avg.</i> degree of creativity, effective transitions, variety and pacing of the visual elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>avg. to below avg.</i> . * Costuming and staging is <i>sometimes</i> creative and provides a <i>less than moderate</i> level of effectiveness. * The inter-relationship between the vocal and musical elements provides an <i>avg. to below avg.</i> level of effectiveness.	Box 3B (71-80) REPertoire EFFECT-Repertoire provides an <i>avg. to above avg.</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>usually</i> entertained with the level of involvement being influenced by an <i>avg. to above avg.</i> degree of creativity, effective transitions, variety and pacing of the visual elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>very good</i> . * Costuming and staging is <i>usually</i> creative and provides a <i>high level</i> of effectiveness. * The inter-relationship between the vocal and musical elements provides an <i>avg. to above avg.</i> level of effectiveness.	Box 4 (81-90) REPertoire EFFECT-Repertoire provides an <i>excellent</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>usually</i> entertained and kept engaged by the creativity, effective transitions, variety and pacing of the visual elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>very good</i> . * Costuming and staging is <i>usually</i> creative and provides a <i>high level</i> of effectiveness. * The inter-relationship between the vocal and musical elements provides a <i>high level</i> of effectiveness.	Box 5 (90-100) REPertoire EFFECT-Repertoire provides an <i>exceptional</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>consistently</i> entertained and kept engaged by the creativity, effective transitions, variety and pacing of the visual elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>excellent</i> . * Costuming and staging is <i>creative</i> and provides a <i>superior</i> level of effectiveness. * The inter-relationship between the vocal and musical elements provides <i>maximum</i> levels of effectiveness.	REPertoire EFFECT (100 pts)
Comments	1 - 2	3 - 4	5 - 6	7	8 - 9	10	TOTAL SCORE (300 pts possible)
						TOTAL SCORE (Possible 300 pts)	

Copyright© 2018, ISSMA, Inc.

Judges Signature \_\_\_\_\_



## ISSMA SHOW CHOIR QUALIFICATION VOCAL PERFORMANCE

	Box 1 (0-30)	Box 2 (31-60)	Box 3A (61-70)	Box 3B (71-80)	Box 4 (81-90)	Box 5 (90-100)	TECHNIQUE (100 pts.)
TECHNIQUE	TECHNIQUE-Performers exhibit <i>very little</i> understanding of good vocal production. * Breath support and control is <i>not</i> apparent. * Tonal focus and projection are <i>not</i> achieved. * Vocal parts are <i>not</i> in tune and well balanced. Individual voices and sections can be heard. * Words are <i>not</i> understood and vowels are seldom uniform. Overall diction is <i>severely</i> lacking. * A majority of the ensemble displays <i>no</i> skill with regard to rhythmic and/or note accuracy. * Ensemble cohesiveness is <i>not</i> achieved.	TECHNIQUE-Performers exhibit a <i>limited</i> understanding of good vocal production. * <i>Frequent</i> lapses in breath support and control occur. * Tonal focus and projection are both <i>severely</i> lacking. * Vocal parts are <i>seldom</i> in tune and well balanced. Individual voices or sections can be heard. * Words are <i>seldom</i> understood and vowels are seldom uniform. Overall diction is <i>severely</i> lacking. * Ensemble displays <i>significant</i> lack of skill with regard to rhythmic and/or note accuracy. * Ensemble cohesiveness is <i>rarely</i> achieved.	TECHNIQUE-Performers exhibit <i>avg. to below avg.</i> control of vocal production. * Breath support and control is <i>avg. to below avg.</i> . * Tonal focus is <i>avg. to below avg.</i> , with inconsistencies as being apparent resulting in hardness and/or breakdowns in projection. * Vocal parts are <i>not</i> always in tune and well balanced. Individual voices or sections may be heard. * Words are <i>frequently</i> not understood and vowels are <i>frequently</i> not uniform. Overall diction is <i>avg. to below avg.</i> . * Ensemble displays <i>avg. to below avg.</i> skill with regard to rhythmic and/or note accuracy. * Cohesiveness may not always be sustained and is <i>avg. to below avg.</i> .	TECHNIQUE-Performers exhibit <i>avg. to above avg.</i> control of vocal production. * Breath support and control is <i>avg. to above avg.</i> . * Tonal focus is <i>good</i> but inconsistencies are apparent resulting in hardness and/or minor breakdowns in projection. * Vocal parts are <i>often</i> in tune and well balanced. Individual voices or sections are <i>sometimes</i> heard. * Words are <i>sometimes</i> understood and vowels are <i>not</i> always uniform. Overall diction is <i>avg. to above avg.</i> . * Ensemble displays <i>avg. to above avg.</i> skill with regard to rhythmic and/or note accuracy with <i>very minor</i> breakdowns in cohesiveness.	TECHNIQUE-Performers exhibit <i>very good</i> control of vocal production. * Minor lapses in breath support and control may occur. * Tonal focus is <i>very good</i> and voices are usually heard. * Vocal parts are <i>generally</i> in tune and well balanced with minor flaws, but with quick recovery. * Words are <i>usually</i> understood and vowels are <i>mostly</i> uniform throughout. * Ensemble displays <i>very good</i> skill with regard to rhythmic and/or note accuracy with <i>very minor</i> breakdowns in cohesiveness.	TECHNIQUE-Performers exhibit <i>superior</i> control of vocal production. * <i>Outstanding</i> breath support and control is maintained throughout. * Tonal focus is <i>never</i> lost and voices are <i>always</i> heard. * Vocal parts are <i>always</i> in tune and well balanced. * Words are <i>clearly</i> understood and vowels are <i>uniform</i> throughout. * Ensemble displays <i>complete</i> control of skill with regard to rhythmic and/or note accuracy with <i>no</i> breakdowns in ensemble cohesiveness.	
ARTISTRY	Box 1 (0-30) ARTISTRY-Performers exhibit <i>very little</i> understanding of the artistry of music performance. * Dynamic contrasts and shaping of phrases are <i>not</i> used to enhance the music. * Vocal production <i>does not</i> reflect an understanding of style or idiomatic interpretation. * Performers <i>do not</i> demonstrate any understanding of professionalism with <i>frequent</i> major lapses occurring in concentration, discipline, and engagement.	Box 2 (31-60) ARTISTRY-Performers exhibit a <i>limited</i> understanding of the artistry of music performance. * Dynamic contrasts and shaping of phrases are <i>seldom</i> used to enhance the music. * Vocal production <i>seldom</i> reflects an understanding of style or idiomatic interpretation. * Performers <i>seldom</i> demonstrate an understanding of professionalism with <i>major</i> lapses occurring in concentration, discipline, and engagement.	Box 3A (61-70) ARTISTRY-Performers exhibit an <i>avg. to below avg.</i> achievement level in the artistry of music performance. * There is an <i>avg. to below avg.</i> use of dynamic contrasts and shaping of phrases to enhance the music. * Vocal production reflects an <i>avg. to below avg.</i> approach to style and idiomatic interpretation. * Performers demonstrate an <i>avg. to below avg.</i> understanding of professionalism with <i>frequent</i> lapses occurring in concentration, discipline, and engagement.	Box 3B (71-80) ARTISTRY-Performers exhibit an <i>avg. to above avg.</i> achievement level in the artistry of music performance. * There is an <i>avg. to above avg.</i> use of dynamic contrasts and shaping of phrases to enhance the music. * Vocal production reflects an <i>avg. to above avg.</i> approach to style and idiomatic interpretation. * Performers demonstrate an <i>avg. to above avg.</i> understanding of professionalism but lapses occur in concentration, discipline, and engagement.	Box 4 (81-90) ARTISTRY-Performers exhibit a <i>high level</i> of achievement in the artistry of music performance. * Dynamics and expressive shaping of phrases are <i>usually</i> used to enhance the communication and emotional aspects of the music. * Vocal production reflects a <i>maturing</i> and <i>sometimes</i> appropriate approach to style and idiomatic interpretation. * A <i>high degree</i> of professionalism is present, but <i>minor</i> lapses in concentration, discipline, and engagement occur.	Box 5 (90-100) ARTISTRY-Performers exhibit <i>superior</i> achievement in the artistry of music performance. * Dynamics and expressive shaping of phrases are <i>used</i> to effectively enhance the communication and emotional aspects of the music. * Vocal production reflects a <i>major</i> and <i>appropriate</i> approach to style and idiomatic interpretation. * The <i>highest</i> degree of professionalism is consistently present.	ARTISTRY (100 pts.)
REPertoire EFFECT	Box 1 (0-30) REPertoire EFFECT-Repertoire provides <i>very little</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>not engaged</i> due to the lack of creativity, transitions, variety and pacing of the vocal elements of the production. * Construction of the repertoire is <i>extremely</i> out of balance with the demonstrated level of performance. * Good balance is <i>not</i> achieved between the vocal ensemble and the accompaniment. * The inter-relationship between the vocal and visual elements is <i>not</i> effective.	Box 2 (31-60) REPertoire EFFECT-Repertoire provides a <i>limited</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>seldom</i> entertained with the level of involvement being influenced by an <i>avg. to below avg.</i> degree of creativity, effective transitions, variety and pacing of the vocal elements of the production. * Construction of the repertoire is <i>seldom</i> well balanced with the demonstrated level of performance. * Good balance is <i>rarely</i> achieved between the vocal ensemble and the accompaniment. * The inter-relationship between the vocal and visual elements provides a <i>limited</i> level of effectiveness.	Box 3A (61-70) REPertoire EFFECT-Repertoire provides an <i>avg. to below avg.</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>seldom</i> entertained with the level of involvement being influenced by an <i>avg. to below avg.</i> degree of creativity, effective transitions, variety and pacing of the vocal elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>avg. to below avg.</i> . * Good balance is <i>between</i> the vocal ensemble and the accompaniment. * The inter-relationship between the vocal and visual elements provides an <i>avg. to below avg.</i> level of effectiveness.	Box 3B (71-80) REPertoire EFFECT-Repertoire provides an <i>avg. to above avg.</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>often</i> entertained with the level of involvement being influenced by an <i>avg. to above avg.</i> degree of creativity, effective transitions, variety and pacing of the vocal elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>avg. to above avg.</i> . * Good balance is <i>usually</i> maintained between the vocal ensemble and the accompaniment. * The inter-relationship between the vocal and visual elements provides an <i>avg. to above avg.</i> level of effectiveness.	Box 4 (81-90) REPertoire EFFECT-Repertoire provides an <i>excellent</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>usually</i> entertained and kept engaged by the creativity, effective transitions, variety and pacing of the vocal elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>very good</i> . * Proper balance is <i>usually</i> maintained between the vocal ensemble and the accompaniment. * The inter-relationship between the vocal and visual elements provides a <i>high level</i> of effectiveness.	Box 5 (90-100) REPertoire EFFECT-Repertoire provides an <i>exceptional</i> opportunity to enhance the effectiveness of the vocal performance. * The audience is <i>consistently</i> entertained and kept engaged by the creativity, effective transitions, variety and pacing of the vocal elements of the production. * Balance between the construction of the repertoire and the demonstrated level of performance is <i>excellent</i> . * Proper balance is <i>maintained</i> between the vocal ensemble and the accompaniment. * The inter-relationship between the vocal and visual elements provides <i>maximum</i> levels of effectiveness.	REPertoire EFFECT (100 pts)
Comments	1 - 2	3 - 4	5 - 6	7	8 - 9	10	TOTAL SCORE (300 pts possible)
						TOTAL SCORE (Possible 300 pts)	

Copyright© 2018, ISSMA, Inc.

Judges Signature \_\_\_\_\_

## APPENDIX E – IRB Approval Letter

IRB-18-32 - Initial: Sacco Committee Letter - Exempt

Office of  
Research Integrity

118 COLLEGE DRIVE #5125 • HATTIESBURG, MS | 601.266.6576 | USM.EDU/ORI



### NOTICE OF INSTITUTIONAL REVIEW BOARD ACTION

The project below has been reviewed by The University of Southern Mississippi Institutional Review Board in accordance with Federal Drug Administration regulations (21 CFR 26, 111), Department of Health and Human Services regulations (45 CFR Part 46), and University Policy to ensure:

- The risks to subjects are minimized and reasonable in relation to the anticipated benefits.
- The selection of subjects is equitable.
- Informed consent is adequate and appropriately documented.
- Where appropriate, the research plan makes adequate provisions for monitoring the data collected to ensure the safety of the subjects.
- Where appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of all data.
- Appropriate additional safeguards have been included to protect vulnerable subjects.
- Any unanticipated, serious, or continuing problems encountered involving risks to subjects must be reported immediately. Problems should be reported to ORI via the Incident template on Cayuse IRB.
- The period of approval is twelve months. An application for renewal must be submitted for projects exceeding twelve months.

PROTOCOL NUMBER: IRB-18-32

PROJECT TITLE: Dissertation: Evaluating Reliability and User-Perceptions of a Showchoir Adjudication Rubric

SCHOOL/PROGRAM: School of Music

RESEARCHER(S): LaDona Tyson, Gregory Fuller

IRB COMMITTEE ACTION: Exempt

CATEGORY: Exempt

Category 2. Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

APPROVED STARTING: February 18, 2019

A handwritten signature in cursive script that reads "Donald Sacco".

Donald Sacco, Ph.D.  
Institutional Review Board Chairperson

## REFERENCES

- American Choral Directors Association. (2020). Contemporary/commercial repertoire. Retrieved January 12, 2020, from [https://acda.org/ACDA/Repertoire/Contemporary\\_Commercial/ACDA/Repertoire/Contemporary\\_Commercial.aspx?hkey=e449108d-7776-4fa9-822a-142521824129](https://acda.org/ACDA/Repertoire/Contemporary_Commercial/ACDA/Repertoire/Contemporary_Commercial.aspx?hkey=e449108d-7776-4fa9-822a-142521824129).
- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57 (5), 13-18. Retrieved from <http://web.b.ebscohost.com.lynx.lib.usm.edu/ehost/pdfviewer/pdfviewer?vid=0&sid=18033328-9a4e-4df5-9f24-795b31593f4c%40pdc-v-sessmgr02>
- Andrade, H.G. (2001). The effects of instructional rubrics on learning to write. *Educational Theory and Practice Faculty Scholarship*. 6. Retrieved from [http://scholarsarchive.library.albany.edu/etap\\_fac\\_scholar/6](http://scholarsarchive.library.albany.edu/etap_fac_scholar/6)
- Andrade, H., Ying, D., & Wang, X. (2008). Putting rubrics to the test: The effect of a model-criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, 27(2), 3-13. doi: 10.1111/j.1745-3992.2008.00118.x
- Arter, J. (2000, April). Rubrics, scoring guides, and performance criteria: Classroom tools for assessing and improving student learning. *Annual Meeting of the American Educational Research Association*. Address conducted at the meeting of American Educational Research Association, New Orleans, LA.
- Asmus, E. P. (1999). Music assessment concepts: A discussion of assessment concepts and models for student assessment introduces this special focus issue. *Music Educators Journal*, 86(2), 19-24. Retrieved from [https://journals.sagepub.com/doi/pdf/10.2307/3399585?casa\\_token=MNB1SM1iGlgAAAAA:SSeydVeBgCXIVoqtipFNNQMPHVGBJ7tiiZbBs1OT\\_VGA5wsLF\\_WNY55KmM1iDytNesK11f7MVyQt3g](https://journals.sagepub.com/doi/pdf/10.2307/3399585?casa_token=MNB1SM1iGlgAAAAA:SSeydVeBgCXIVoqtipFNNQMPHVGBJ7tiiZbBs1OT_VGA5wsLF_WNY55KmM1iDytNesK11f7MVyQt3g)
- Austin, J.R. (1988). The effect of music contest format on self-concept, motivation, achievement, and attitude of elementary band students. *The Journal of Research in Music Education*, 36(2), 95-107. Retrieved from <http://www.jstor.org/stable/3345243>
- Austin, J.R. (1990). Is music education the loser?. *Music Educators Journal*, 76(6), 21-25. Retrieved from <http://www.jstor.org/stable/3400964>
- Bargainnier, S. (2003). Fundamentals of rubrics. *Pacific Crest*, 1-4. Retrieved from [https://www.webpages.uidaho.edu/ele/scholars/Practices/Evaluating\\_Projects/Resources/Using\\_Rubrics.pdf](https://www.webpages.uidaho.edu/ele/scholars/Practices/Evaluating_Projects/Resources/Using_Rubrics.pdf)



- Barry, N. (2009). Evaluating music performance: Politics, pitfalls, and successful practices. *College Music Symposium*, 49/50, 246-256. Retrieved from <http://www.jstor.org/stable/41225250>
- Bergee, M. J. (2003). Faculty interjudge reliability of music performance evaluation. *Journal of Research in Music Education*, 51(2), 137-150. <https://doi.org/10.2307/3345847>
- 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), 342-353. Retrieved from [https://journals.sagepub.com/doi/pdf/10.2307/3345660?casa\\_token=v1x4Jv4D4dAAAAAA:vX\\_BNleOuQ9Kdmf5FRNMzoraKk1qF73mmM0gipwe2PWvOmGV oC\\_uJerzRn7A\\_6prtBNd0DF0aJs](https://journals.sagepub.com/doi/pdf/10.2307/3345660?casa_token=v1x4Jv4D4dAAAAAA:vX_BNleOuQ9Kdmf5FRNMzoraKk1qF73mmM0gipwe2PWvOmGV oC_uJerzRn7A_6prtBNd0DF0aJs)
- Boyle, J. D., & Radocy, R. E. (1987). *Measurement and evaluation of musical experiences*. New York, NY: Schirmer books.
- Brakel, T. D. (2006). Inter-judge reliability of the Indiana State School Music Association high school instrumental festival. *Journal of Band Research*, 42(1), 59-69. Retrieved from <https://search.proquest.com/docview/1092044?pq-origsite=gscholar>
- Brodie, L., & Gibbings, P. (2009). Comparison of PBL assessment rubrics. In *Proceedings of the Research in Engineering Education Symposium (REES 2009)* (pp. 1-9). University of Melbourne. Retrieved from [https://eprints.usq.edu.au/5480/2/Brodie\\_Gibbings\\_REES\\_2009\\_PV.pdf](https://eprints.usq.edu.au/5480/2/Brodie_Gibbings_REES_2009_PV.pdf)
- Brookhart, S. M. (2018). Appropriate criteria: key to effective rubrics. *Frontiers in Education*, 3, p. 22. Retrieved from <https://www.frontiersin.org/articles/10.3389/feduc.2018.00022/full>
- Bumsed, V., Hinkle, D., & King, S. (1985). Performance evaluation reliability at selected concert festivals. *Journal of Band Research*, 21(1), 22-29. Retrieved from <http://lynx.lib.usm.edu/docview/1096689?accountid=13946>
- Burnsed, V., & Sochinski, J. (1983). Research on competitions: Surveys reveal how students, parents, directors, and administrators feel about competitions. *Music Educators Journal*, 70(2), 25-27. <https://doi.org/10.2307/3401207>
- Burrack, F. W., Payne, P., Bazan, D. E., & Hellman, D. S. (2014). The impact of budget cutbacks on music teaching positions and district funding in three Midwestern states. *Update: Applications of Research in Music Education*, 33(1), 36-41. <https://doi.org/10.1177/8755123314521039>

- Carlin, N., Rozmus, C., Spike, J., Willcockson, I., Seifert, W., Chappell, C., ... & Lunstroth, R. (2011). The health professional ethics rubric: Practical assessment in ethics education for health professional schools. *Journal of Academic Ethics*, 9(4), 277. <https://doi.org/10.1007/s10805-011->
- Cassidy, J.W., & Sims, W.L. (1991). Effects of special education labels on peers' and adults' evaluations of a handicapped youth choir. *Journal of Research in Music Education*, 39 (1), 23-34. Retrieved from [https://www.jstor.org/stable/3344606?seq=1&cid=pdf-reference#references\\_tab\\_contents](https://www.jstor.org/stable/3344606?seq=1&cid=pdf-reference#references_tab_contents)"
- 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. <https://doi.org/10.1177/0022429409333405>
- East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing*, 14(2), 88-115. Retrieved from <https://doi.org/10.1016/j.asw.2009.04.001>
- Eisner, E. (2007). Interlude: Assessment and evaluation in education and the arts. In Bresler, L. (Ed.), *International Handbook of Research in Arts Education* (423-426). The Netherlands: Springer.
- FAME Show Choir National Championship Series. (n.d.) Retrieved from [www.showchoirs.org](http://www.showchoirs.org).
- Fiske Jr, H. E. (1975). Judge-group differences in the rating of secondary school trumpet performances. *Journal of Research in Music Education*, 23(3), 186-196. <https://doi.org/10.2307/3344643>
- Fiske, H. (1977). Who's to judge? New insights into performance evaluation. *Music Educators Journal*, 64(4), 22-25. Retrieved from [www.jstor.org/stable/3395371](http://www.jstor.org/stable/3395371)
- Fiske, H. E. (1978). The effect of a training procedure in musical performance evaluation on judge reliability. *Ontario Educational Research Council Report*.
- Forbes, G. W. (1994). Evaluative music festivals and contests—Are they fair?. *Update: Applications of Research in Music Education*, 12(2), 16-20. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.844.8069&rep=rep1&type=pdf>

- Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vols. 1-4). Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781506326139
- Gariepy, L. W., Stout, J. A., & Hodge, M. L. (2016). Using rubrics to assess learning in course-integrated library instruction. *Portal: Libraries and the Academy*, 16(3), 491–509. Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ1106890&site=ehost-live>
- Gerritson, M. (2013). *Rubrics as a mitigating instrument for bias in the grading of student writing* (Doctoral dissertation) Retrieved from ProQuest. (3566384)
- Gipps, C. (2011). *Beyond testing (classic edition): Towards a theory of educational assessment*. New York: Routledge.
- Green, D. (1984). Choralography: Expressive movement for choral singing. *The Choral Journal*, 25(3), 19-23. Retrieved from [www.jstor.org/stable/23546460](http://www.jstor.org/stable/23546460)
- Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education*, 25(12), 1509-1528. doi:10.1080/0950069022000038268
- Hash, P. (2012). An analysis of the ratings and interrater reliability of high school band contests. *Journal of Research in Music Education*, 60(1), 81-100. Retrieved from <http://www.jstor.org.lynx.lib.usm.edu/stable/41348853>
- Hash, P. (2013). Large-group contest ratings and music teacher evaluation: Issues and recommendations. *Arts Education Policy Review*, 114(4), 163–169. <https://doi-org.lynx.lib.usm.edu/10.1080/10632913.2013.826035>
- Heart of America. (n.d.) Retrieved from <https://www.hoaentertainment.com/>.
- Herro, D., Quigley, C., Andrews, J., & Delacruz, G. (2017). Co-Measure: Developing an Assessment for Student Collaboration in STEAM Activities. *International Journal of STEM Education*, 4. Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ1181927&site=ehost-live>
- Hung, H.-T., Chiu, Y.-C. J., & Yeh, H.-C. (2013). Multimodal assessment of and for learning: A Theory-driven design rubric. *British Journal of Educational Technology*, 44(3), 400–409. Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ1010158&site=ehost-live>

- Indiana State School Music Association. (2020). Retrieved January 7, 2020, from <https://www.issma.net/adjudicator.php>.
- Indiana State School Music Association. (n.d.). Become a certified ISSMA adjudicator. Retrieved from [https://issmaacademy.instructure.com/courses/15/pages/become-a-certified-issma-adjudicator?module\\_item\\_id=242](https://issmaacademy.instructure.com/courses/15/pages/become-a-certified-issma-adjudicator?module_item_id=242)
- Field, A. (2005) Intraclass correlation. In B. Everitt editor & D. Howell (eds.), *Encyclopedia of statistics in behavioral science* (948-954). Chichester: John Wiley & Sons. <https://doi-org.lynx.lib.usm.edu/10.1002/0470013192.bsa313>
- Iowa High School Music Association. (n.d.) Sanctioning information. Retrieved from <https://ihsma.org/sanctioning-information/>.
- Johnson, E.S., Moylan, L.A., Crawford, A., & Zheng, Y. (2019). Developing a comprehensive instruction observation rubric for special education teachers. *Reading and Writing Quarterly*, 35(2), 118-136. Retrieved from <https://files.eric.ed.gov/fulltext/ED598440.pdf>
- Johnson, S. M. (1978). Evaluation in Continuing Education: The Critical Feedback Loop. *American Journal of Pharmaceutical Education*, 42(4), 396-401. Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ194432&site=ehost-live>
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144. Retrieved from <https://doi.org/10.1016/j.edurev.2007.05.002>
- King, S. E., & Burnsed, V. (2009). A study of the reliability of adjudicator ratings at the 2005 Virginia Band and Orchestra Directors Association state marching band festivals. *Journal of Band Research*, 45(1), 27-32, 62-63. Retrieved from <http://lynx.lib.usm.edu/docview/216236038?accountid=13946>
- Koo TK, Li MY (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*. 15(2), 155–63. doi:10.1016/j.jcm.2016.02.012. PMC 4913118. PMID 27330520.
- Kural, F. (2018). Does indirect writing assessment have any relevance to direct writing assessment? Focus on validity and reliability. *Journal of Language and Linguistic Studies*, 14(4), 342-351. Retrieved from <http://web.b.ebscohost.com.lynx.lib.usm.edu/ehost/pdfviewer/pdfviewer?vid=0&sid=eeabc8c0-cd30-413e-95bb-56d9bf5b5fa1%40pdc-v-sessmgr05>

- Latimer Jr, M. E. (2013). Choral performance assessment: An overview of research to date. *The Phenomenon of Singing*, 7, 81-91. Retrieved from <http://journals.library.mun.ca/ojs/index.php/singing/article/viewFile/901/779>
- 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, 67-82. Retrieved from <http://lynx.lib.usm.edu/docview/222276253?accountid=13946>
- Latimer, 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. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1177/0022429410369836>
- Lee, E & Lee, S. (2009). Effects of instructional rubrics on class engagement behaviors and the achievement of lesson objectives by students with mild mental retardation and their typical peers. *Education and Training in Developmental Disabilities*, 44(3), 396-408. Retrieved from <https://www.jstor.org/stable/24233483>
- Liao, M. Y. (2002). *The effects of gesture and movement training on the intonation and tone quality of children's choral singing* (Doctoral dissertation). Retrieved from <http://etheses.whiterose.ac.uk/12831/>. (12831)
- Logan, K., & Mountain, L. (2018). Writing instruction in chemistry classes: Developing prompts and rubrics. *Journal of Chemical Education*, 95(10), 1692–1700. Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ1195647&site=ehost-live>
- McKenna, S. (2007). Assessment rubrics as a means of making academic expectations explicit. *The Independent Journal of Teaching and Learning*, 2(1), 22-30. Retrieved from [http://journals.co.za/docserver/fulIext/jitl/2/1/jitl\\_v2\\_a4.pdf?expires=1521040462&id=id&accname=guest&checksum=1CADE525CFFE06DC6F34204E80E80E90](http://journals.co.za/docserver/fulIext/jitl/2/1/jitl_v2_a4.pdf?expires=1521040462&id=id&accname=guest&checksum=1CADE525CFFE06DC6F34204E80E80E90)
- McPherson, G. E., & Schubert, E. (2004). Measuring performance enhancement in music. In A. Williamon (ed.), *Musical excellence: Strategies and techniques to enhance performance* (61-82). London: Oxford University Press.
- McPherson, G. E., & Thompson, W. F. (1998). Assessing music performance: Issues and influences. *Research Studies in Music Education*, 10(1), 12-24. Retrieved from [https://www.researchgate.net/profile/Gary\\_Mcpherson2/publication/249831078\\_Assessing\\_Music\\_Performance\\_Issues\\_and\\_Influences/links/0deec530333ad29a52000000/Assessing-Music-Performance-Issues-and-Influences.pdf](https://www.researchgate.net/profile/Gary_Mcpherson2/publication/249831078_Assessing_Music_Performance_Issues_and_Influences/links/0deec530333ad29a52000000/Assessing-Music-Performance-Issues-and-Influences.pdf)

- Milanowski, A. (2017). Lower performance evaluation practice ratings for teachers of disadvantaged students: Bias or reflection of reality?. *AERA Open*, 3(1). Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ1194177&site=ehost-live>
- Millard, E. (2014). *The role of ensemble competitions in choral music education* (Doctoral Dissertation). Retrieved from ProQuest. (Order No. 3638044)
- Mississippi Music Educators Association. (2020). Retrieved from <https://msmea.org/about-us/choral-adjudicators-course>.
- Moser, C. (1951). Interview bias. *Review of the International Statistical Institute*, 19(1), 28-40. doi:10.2307/1401500
- Murphy, R. (2007). Harmonizing assessment and music in the classroom. In L. Bresler (Ed.), *International Handbook of Research in Arts Education* (361-379). The Netherlands: Springer.
- Napoles, J. (2009). The effects of score use on musicians' ratings of choral performances. *Journal of Research in Music Education*, 57(3), 267-279. <https://doi.org/10.1177/0022429409343423>
- 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.
- Rawlings, J. R. (2019). Benefits and challenges of large-ensemble instrumental music adjudicated events: Insights from experienced music teachers. *Update: Applications of Research in Music Education*, 37(2), 46-53. <https://doi.org/10.1177/8755123318777824>
- Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing*, 15(1), 18-39. <https://doi.org/10.1016/j.asw.2010.01.003>
- Rogers, G. L. (1982). *Attitudes of high school band directors, band members, parents, and principals toward marching band contests* (Doctoral Dissertation). Retrieved from ProQuest. (8306179)
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education: principles, policy & practice*, 5(1), 77-84. Retrieved from <http://proquest.umi.com/pqdlink?did=30132759&sid=1&Fmt=3&clientId=15814&RQT=309&VName=PQD>

- Schmalstieg, E. (1972). The construction of a reliable test for assessing musical performance. *Journal of Research in Music Education*, 20(2), 280-282. Retrieved from [www.jstor.org/stable/3344095](http://www.jstor.org/stable/3344095)
- Schmidt, C. (2005). Relations among motivation, performance achievement, and music experience variables in secondary instrumental music students. *Journal of Research in Music Education*, 53(2), 134-147. Retrieved from [www.jstor.org/stable/3345514](http://www.jstor.org/stable/3345514)
- Shaw, K., & Thomas, K. (2005). Show choirs: Competitive show choir festivals: What are the benefits?. *The Choral Journal*, 45(7), 107-109. Retrieved from <https://www.jstor.org/stable/23555351>
- Shipman, D., Roa, M., Hooten, J., & Wang, Z. J. (2012). Using the analytic rubric as an evaluation tool in nursing education: The positive and the negative. *Nurse Education Today*, 32(3), 246-249. Retrieved from (DOI): 10.1016/j.nedt.2011.04.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420-428. doi: 10.1037/0033-2909.86.2.420
- Silvestri, L., & Oescher, J. (2006). Using rubrics to increase the reliability of assessment in health classes. *International Electronic Journal of Health Education*, 9, 925-30. Retrieved from <http://member.aahperd.org/iejhe/template.cfm?template=currentIssue.cfm#volume9>
- Show Choir Community. (2002-2020). Retrieved from <https://www.showchoir.com/choirs/>
- Sparrow, S. M. (2004). Describing the ball: Improve teaching by using rubrics-explicit grading criteria. *Mich. St. L. Rev.*, 1. Retrieved from [http://scholars.unh.edu/cgi/viewcontent.cgi?article=1031&context=law\\_facpub](http://scholars.unh.edu/cgi/viewcontent.cgi?article=1031&context=law_facpub)
- Spohn, C. (2008). Teacher perspectives on No Child Left Behind and arts education: A case study. *Arts Education Policy Review*, 109(4), 3-12. <https://doi.org/10.3200/AEPR.109.4.3-12>
- Stanford, P., & Reeves, S. (2005). Assessment that drives instruction. *Teaching Exceptional Children*, 37(4), 18-22. Retrieved from <http://journals.sagepub.com/doi/pdf/10.1177/004005990503700402>

- Stanley, M., Brooker, R., & Gilbert, R. (2002). Examiner perceptions of using criteria in music performance assessment. *Research Studies in Music Education*, 18, 46–56. doi:10.1177/1 321103X020180010601
- Stegman, S.F. (2009). Michigan state adjudicated choral festivals: Revising the adjudication process. *Music Educators Journal*, 95(4), 62-65. doi: 10.1177/0027432109335852
- Stellmack, M. A., Konheim-Kalkstein, Y. L., Manor, J. E., Massey, A. R., & Schmitz, J. P. (2009). An assessment of reliability and validity of a rubric for grading apa-style introductions. *Teaching of Psychology*, 36(2), 102-107. Retrieved from <http://www.informaworld.com.lynx.lib.usm.edu/openurl?genre=article&id=doi:10.1080/00986280902739776>
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4), 1-19. Retrieved from <http://pareonline.net/GETVN.ASP?V=9&N=4>
- Sundeen, T.H. (2014) Instructional rubrics: Effects of presentation options on writing quality. *Assessing Writing*, 21(3), 74-88. Retrieved from <https://doi.org/10.1016/j.asw.2014.03.003>
- Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a ‘universal’ rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, 36(5), 509-547. doi:10.1080/02602930903540991
- Tractenberg, R. E., Umans, J. G., & McCarter, R. J. (2010). A mastery rubric: Guiding curriculum design, admissions and development of course objectives. *Assessment & Evaluation in Higher Education*, 35(1), 17–35. Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ879965&site=ehost-live>
- Weaver, M. and Hart, C. (2011). *Sweat, tears, and jazz hands: The official history of show choir from vaudeville to glee*. Milwaukee: Hal Leonard Books.
- Wesolowski, B. C. (2012). Understanding and developing rubrics for music performance assessment. *Music Educators Journal*, 98(3), 36-42. Retrieved from <https://doi.org/10.1177/0027432111432524>
- Wesolowski, B.C., Wind, S. A., & Engelhard, G. (2015). Rater fairness in music performance assessment: Evaluating model-data fit and differential rater functioning. *Musicae Scientiae*, 19(2) 147-170. doi:10.1177/1029864915589014.



- Wesolowski, B. C., Athanas, M. I., Burton, J. S., Edwards, A. S., Edwards, K. E., Goins, Q. R., ... Thompson, J. E. (2018). Judgmental standard setting: The development of objective content and performance standards for secondary-level solo instrumental music assessment. *Journal of Research in Music Education*, 66(2), 224–245. Retrieved from <http://search.ebscohost.com.lynx.lib.usm.edu/login.aspx?direct=true&db=eric&AN=EJ1180961&site=ehost-live>
- West, C. (2012). Teaching music in an era of high-stakes testing and budget reductions. *Arts Education Policy Review*, 113(2), 75-79. <https://doi.org/10.1080/10632913.2012.656503>
- Wherry Sr, R. J., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology*, 35(3), 521-551. doi:10.1111/j.1744-6570.1982.tb02208.x
- Wilson, M. (2006). *Rethinking rubrics in writing assessment*. Portsmouth: Heinemann.
- WorldStrides Educational Travel and Experiences. (2020). Heritage festivals. Retrieved from <https://worldstrides.com/collections/heritage-festivals/>.