

ABSTRACT

Title of Thesis: ASSESSMENT IN SECONDARY MUSIC CLASSROOMS: A REPLICATION STUDY

William Gilbert Gonzales, Masters of Arts, 2017

Thesis Directed By: Dr. Kenneth Elpus, Music Education

The purpose of this study was to examine assessment practices of secondary school music teachers in the mid-Atlantic region of the United States by replicating the research methods of Russell & Austin (2010). A total of 4,083 invitation emails were sent once a week over the course of three weeks, 840 emails were reported opened by recipients. Though a total 291 responses were received, only 185 questionnaires were usable for analysis. As in Russell & Austin (2010), the majority of respondents reported using traditional letter grades and that their classes are weighted equally with other general education classes for calculating students' overall grade point average (GPA) and count towards graduation requirements. Participants used both achievement and non-achievement criteria for determining students' grades. However, on average, the aggregate of non-achievement criteria was weightily slightly heavier than achievement criteria. Regarding influence of respondents specific teaching contexts on grading methods, significant influences were found for teaching specialization, assessment confidence, and instructional time.

ASSESSMENT IN SECONDARY MUSIC CLASSROOMS:
A REPLICATION STUDY

by

William Gilbert Gonzales

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Masters of Arts
2017

Advisory Committee:
Professor Kenneth Elpus, Chair
Professor Stephanie Prichard
Professor Michael Hewitt

© Copyright by
William Gilbert Gonzales

2017

Acknowledgements

I would like to thank all of those who have offered me guidance and encouragement throughout my tenure as a graduate student and especially during the process of completing this thesis.

First, Dr. Kenneth Elpus, thank you for your direction and flexibility as I worked on this project as a part-time student. I greatly appreciated your willingness to advise over the phone on Sunday evenings or through text message while you worked a chorus festival. You have been a great inspiration to me throughout my time at the University of Maryland.

To Dr. Stephanie Prichard and Dr. Michael Hewitt, thank you for sitting on my thesis committee and providing such valuable feedback and suggestions.

Thank you, Dr. Joshua Russell of the University of Hartford, for granting me permission to use the Secondary School Music Assessment Questionnaire for this project.

Special thanks to my colleagues at Strathmore Children's Chorus and Newport Mill Middle School for your patience and support as I burned the candle at both ends. Also, I would like to thank my students at both institutions who were some of my biggest cheerleaders during this process.

To my family, all of whom provided me tremendous strength all the way from the New Mexico and Arizona. I particularly want to thank my father, Sam Gonzales, for his unwavering encouragement and my mother, Sandra White, for the constant inspiration.

Lastly I thank my husband, Derek Leininger, for his continued support and occasional edits.

Table of Contents

Acknowledgements.....	ii
Table of Contents.....	iii
List of Tables.....	v
Statement of the Problem.....	1
Purpose of Study.....	3
Research Questions.....	3
Need for Study.....	4
Definitions.....	5
Limitations of Study.....	5
Organization of Study.....	6
Review of Literature.....	8
Assessment in US General Education and Music Education.....	8
Recommended Assessment Practices in Music Education.....	11
Perceptions of Assessment in Music Education.....	14
Grading Using Non-Achievement Criteria.....	17
Assessment Practices of In-Service Music Teachers.....	21
Method.....	28
Replication Guidelines.....	28
Instrument.....	29
Participants.....	31
Procedure.....	34
Results.....	36
Descriptive Statistics.....	36
Inferential Analyses.....	45
Discussion.....	49
Research Question 1.....	50
Research Question 2.....	51
Research Question 3.....	53
Comparison Between Original and Replication Studies.....	55

Implications and Conclusion.....	56
Secondary School Music Assessment Questionnaire	59
References.....	69

List of Tables

Table 3.1: Distribution of Sample by State Compared to Distribution of Mid-Atlantic Active NAFME Members as of January 2017	33
Table 4.1: Assessment Criteria and Weights Used in Grading Policies	39
Table 4.2: Factor Considered by Music Teachers When Assessing Attendance and Attitude	40
Table 4.3: Percentage of Teachers Using Specific Objectives and Formats for Written Assessment	41
Table 4.4: Percentage of Teachers Using Specific Objectives and Formats for Performance Assessments	42
Table 4.5: Multivariate and Univariate Analyses for Significant Relations Between Teaching Level, Teaching Specialization, and Assessment Criteria	45
Table 4.6: Correlation Between Assessment Criteria Weight and Teaching Contexts	46

Chapter 1

Statement of the Problem

Secondary music classes in the United States predominantly function on a performance-based ensemble model, such as band, chorus, or orchestra (Abril & Gault, 2008; Elpus & Abril, 2011). These classrooms have historically mirrored the environment and format of conservatory or extracurricular ensemble rehearsals, wherein the teacher assumes the role of the conductor and students are the musicians. Lessons typically consist of a tuning/warm-up and then repetitious cycling through of various repertoire pieces. Student musicians play music selected by the teacher/conductor and rely on his/her response to make improvements to their individual and collective performance. This classroom model has been pervasive within the field of music education since the emergence of music ensembles in the American public school curriculum and one can reasonably assume it will continue as such for the foreseeable future. Within the past two decades, however, assessment practices and specifically a focus on reliable measurement of student achievement has permeated the national discussion in American education. Consequently, music education researchers have begun to consider how these concepts and methods apply to music classrooms.

Reliable assessment practices and valid student achievement data is now a standard facet of American public education. This accountability trend is a direct result of national education policy from the early 2000s, such as the No Child Left Behind Act (NCLB), but its roots can be traced to federal incentives from the mid-twentieth century. Though the emphasis on high-stakes standardized testing are commonly perceived as having negative effects on enrollment in secondary music courses (Abril & Gault, 2008),

the US educational policy driving the accountability trend has been shown to have little to no effect on the overall number of middle and high school student enrolled in music (Elpus, 2014). Considering these observations, the true impact of assessment policy on music education can be seen in the push for more reliable grading practices and improved documentation of student achievement.

Under the performance-based large ensemble model, music teachers devote a high majority of instruction time to group rehearsal. Prioritization of ensemble rehearsal, though conducive to performance objectives, limits time for individualized written or authentic assessment and requires music teachers to determine students' grades through informal observation. This process is similar to participation grading, where student scores are based on the vague description of a complex and multifaceted performance tasks. Teachers using this method likely ascribe each student's grade retroactively, as the act of conducting rehearsal doesn't allow for real-time, individualized student record keeping and relying instead on subjective recollection of student participation. Additionally, music teachers often rely on student metrics unrelated to their understanding of academic content such as daily attendance, concert attendance, and rehearsal behavior or etiquette. It should be mentioned these forms of grading are not exclusive to music education, nor are they the only forms of assessment in the field. Despite this, studies that sought to audit the various assessment practices of secondary music classrooms found, to date, that informal observation assessment and student attendance and behavior continue to be determining factors of students' overall music grade (Russell & Austin, 2010). Considering these findings based on regional data from the US Southwest, the present study seeks to examine the grading practices of music

teachers in a separate US region in order gain a clearer understanding of assessment in secondary music education.

Purpose of Study

The purpose of this replication study is to examine the assessment practices of secondary public school music teachers in the mid-Atlantic region of the United States. This project will reproduce the research methods conducted in Russell & Austin (2010), using their instrument and models of statistical analysis, in an effort to answer the same research questions regarding assessment practices of secondary music teachers.

Comparing the assessment practice results of this project with music teacher demographics and music program context will yield insights as to the effectiveness of collegiate assessment instruction, school administration guidelines, and continued professional development. Additionally, this study will explore the possibility of trends in assessment among secondary music teachers in the Mid-Atlantic region of the United States and to determine if assessment practices and philosophies are correlated with specific teaching contexts and demographics. Though the direct outcomes of this project merely provide an audit of regionalized music assessment practices and philosophies, this study will contribute to music education scholarship in the hope to improve assessment practices by secondary music teachers.

Research Questions

This study's research questions are an exact replication of those from Russell & Austin (2010) applied secondary music teachers of the Mid-Atlantic region of the United States:

1. What types of school district frameworks and classroom contexts are secondary music teachers operating within as they assess learning and grade students?
2. Which specific assessment and grading practices are most commonly employed by secondary music teachers?
3. Do any contextual or individual difference variables influence secondary music teachers' assessment and grading practices?

Need for Study

The literature review revealed fewer than five studies over the past two decades focused specifically on the assessment practices of secondary music teachers. The studies have provided an empirical representation of the various forms of grading practices in public school music education classrooms. All studies on music assessment practices indicated a wide use of assessment practices which rely primarily on student attendance and attitude rather than performance ability or music knowledge. These practices are consistently used by secondary music teachers despite years of teaching experience, education, or individual teaching contexts or demographics (Kotora, 2005; McCoy 1988, 1991; Pierre & Wuttke, 2015; Russell & Austin, 2010). Additionally, the researchers from the aforementioned studies called for continued scholarship on the topic of music assessment practices.

Periodically glimpsing into the classroom to observe how student progress is being recorded can benefit teachers and education researchers alike. This need is intensified by the accountability trend in US public education and the ongoing advocacy for music education as a curricular fixture of the overall educational experience.

The review of the literature also presented a dearth of replication studies in music educational scholarship as well as the larger field of social science in whole. Thus, in addition to expanding the findings of Russell & Austin (2010), this project will work to diversify the field music education research by offering a replication study format.

Definitions

The following section provides definitions for common terms and concepts used in the literature of assessment (Asmus, 1999; Fautley & Colwell, 2012; Reynolds, et al., 2010; Russell & Austin, 2010) and current research.

Assessment: The process of gathering data to support a claim or understanding about the person(s) or objects from which data was gathered.

Test: The tool used to gather data for assessment. Tests can be designed and used in a variety of ways as to measure different qualities of subject output.

Achievement criteria: Grading standards that represent a desired level of understanding or the acquisition of knowledge pertaining to a specific content or objective.

Non-achievement criteria: Metrics unrelated to academic achievement or course objectives based on student attendance, participation, and attitude or behavior in class.

Limitations of Study

While the main objective of this study is to gather data on secondary music teachers' grading practices, their teaching experience, and their classroom context in order to improve assessment in music education, the reach and implications of the research face various limitations. This study is contained to the Mid-Atlantic region of

the United States and respondents' data may not represent the larger national picture of music assessment in the United States. A low response rate may be an indicator of unfamiliarity or insecurity with assessment by secondary music teachers in this area. Additionally, and more crucially, I acknowledge that quality music education is not a direct result of excessive testing or in-depth knowledge of recommended assessment practices. Nor does the use of non-achievement student data necessarily imply poor music instruction. The assessment methods written about in this study are mere suggestions for maintaining reliable data regarding student growth and understanding for the benefit of the student and the educator.

The most significant limitation to this study surfaced at the close of the three-week data collection window. The email method used to distribute the questionnaire, which included one initial initiation email and two subsequent reminder emails, produced a response rate between 5 - 22%. Creswell (2002) references 50% responses rate as a preferred standard among leading educational journals and allows for decrease in bias reporting of a population (p. 390). Considering this project's indiscernible response rate and possible response-bias, the findings, analyses, and implications developed in the proceeding chapters will be reported as representing the participating teachers and not generalized to the larger population of Mid-Atlantic secondary music teachers as a whole.

Organization of Study

This study is organized into five chapters, the first of which being the statement of the problem regarding assessment in secondary music education. Chapter two addresses previous literature dealing with the larger topic of assessment in both music and general education. Chapter three details the methods used in conducting this replication study

modeled after Russell & Austin (2010) and the sample of music teachers in the Mid-Atlantic region of the United States. The results of the survey and its subsequent statistical analyses are presented in chapter four. Finally, chapter five discusses the implications of the findings, comparisons to the original study Russell & Austin (2010), reflections on the process, and recommendations for future research regarding assessment in secondary music education.

Chapter 2

Review of Literature

Assessment in US General Education and Music Education

Assessing student progress and achievement is an essential step of the educational process. The success of teachers, students, and the institution of public education alike is dependent upon the data (and interpreted meaning) gleaned from student assessment. Though general concepts regarding assessment are well established in the educational community, the use and role of assessment in US public education is not entirely unified across institution levels. Teachers and community members criticize schools' and districts' over reliance and prioritization of assessment results by highlighting the negative affect this focus has on the classroom and students. The bodies pushing for more accountability and standardization depict this frustration and resistance to assessment as a product of misunderstanding. The dissociation of various forms of assessment and their implementation in education systems across the United States are also blamed. Despite the ongoing debate, it is beneficial to highlight American educational history and the national policies that have led to the emphasis on assessment as federal and state level priorities have lasting ripple effects regarding accountability at both the school and classroom levels. This chapter will provide an examination of assessment both in general and music education by addressing policy, perception, and application. The chapter focuses particularly on practices in secondary music classrooms and the ill-advised use of non-achievement criteria for determining student grades.

United States education is currently structured around an evolving accountability and assessment trend. The focus on assessment was strongly intensified by No Child Left

Behind (NCLB), the 2002 educational legislation which reauthorized the 1960's Elementary and Secondary Education Act and required the implementation of annual standardized testing for students with high-stakes consequences for schools not meeting predetermined progress. As of late 2015, the United States congress signed into law new educational legislation, the Every Student Succeeds Act (ESSA), which made considerable changes to the widely unpopular NCLB, including transferring most of the federal oversight of assessment to the state level. Despite the bipartisan praise of ESSA for revising and improving NCLB, the legislation maintains a heavy focus on accountability through standardized testing and student assessment. Music as an academic subject area does not have a federally mandated standardized test, however the influence of the accountability movement has reached the music classroom as US music teachers are expected to conduct reliable assessments to determine student achievement.

According to recent observations, assessment in American secondary music education has been predominantly formative assessment. Such a form of assessment takes place throughout the learning process and informs instruction (Fautley & Colwell, 2012; Abeles, 2010). This is separate from summative assessment, where measurement is conducted after an instructional period and used to determine if students achieve a specific educational objective. In ensemble classes, music teachers often organically conduct formative assessment throughout the rehearsal process by providing evaluative oral feedback to students' collective or individual performance of music. However, Fautley & Colwell (2012) acknowledge confusion among American music teachers regarding the difference between the formative assessment and summative, as "the formative use of summative assessment" often occurs in music classes, where the

implemented assessment is ineffective in providing students the opportunity to improve (p. 483).

Music teachers have been advised to adapt their formative assessment practices to provide clear expectations and tangible results of achievement. Abeles (2010) recommends alternative approaches of music assessment that focus on testing and documenting students' achievement of observable musical tasks as opposed to the default traditional written assignments which focuses on students' low-level cognitive understanding of music knowledge (p. 176). At the center of all alternative approaches is the concept of authentic assessment, which measures students' performance of tasks or objectives with real world implications. To reliably conduct authentic assessment teachers must implement rubrics, checklists, and other record-keeping methods for evaluating performance. Additionally, student journals or portfolios are recommended for evaluating students' higher-level comprehensive understanding of music beyond performance ability (Abeles, 2010; Asmus, 1999; Fautley & Colwell, 2012). The various recommended assessment improvements prioritize reliability and validity, which suggests the practice of teaching music is following the national trend of accountability and standardization in education.

The National Association for Music Educators (NAfME, 2016) is currently piloting the Model Cornerstone Assessments (MCAs), which consist of guidelines and supporting documentation for assessing recommended performance tasks. These assessments evaluate students' performance in three separate areas of the artistic processes (drawn from the National Core Arts Standards), i.e. creating, performing, responding (NAfME, 2016). The guidelines and documents can be customized to fit the

music teachers' assessment needs, yet still provide teachers and students the means to administer and track student musical progress and achievement via a standardized approach. NAFME states that, if conducted with integrity, the MCA provide music teachers "valid assessments of student learning [that] can reliably document student growth throughout a music program" (NAFME, 2016).

Considering the ubiquitous and highly consequential emphasis on reliable assessment in American general and music education, many researchers have found it poignant to study the assessment literacy of teachers within this framework, that is based on assessment practices in use, assessment philosophy, and perceptions of assessment preparation. Deluca and Bellara (2013) define assessment literacy as educators who are able to present self-created assessments that are valid and reliable. Additionally, the authors argue teachers should be competent in ascertaining levels of student learning and be able to both communicate meaning and tailor instruction based on their findings (p. 1). Though Deluca and Bellara focus primarily on teachers of general education courses (English, math, science, history), the same standard is expected of music teachers at both primary and secondary levels.

Recommended Assessment Practices in Music Education

In addition to chapters written in scholarly handbooks and the guidelines and standards presented by NAFME and the National Coalition of Core Arts Standards, many scholars and practitioners have published articles and books which outline and guide music teachers' implementation of reliable and valid assessments in the secondary music classroom. I henceforth will refer to these publications as music educator practitioners' texts.

Asmus (1999) provides music teachers an introduction to constructing accountable assessment. The article first addresses resistance among music educators to use the common objective of assessment on the grounds that music is a subjective aesthetic endeavor. The author suggests reliable assessment of art is possible if the teacher takes necessary steps to clearly deconstruct vague and abstract concepts, such as “musicality,” into a series of clearly defined and achievable behaviors or tasks (p. 22). Asmus provides practitioners with a glossary of assessment terms, and practical suggestions intended to increase assessment reliability and validity. The text offered a philosophical justification for incorporating assessment into the music classroom at a time when teachers were skeptical about assessment in the arts and the possibility of changes to the field of music education resulting from its implementation. Asmus (1999) appears early in the scholarship on assessment in music education and its introductory-level content suggests a deficit of knowledge among music teachers on this topic. The literature has gradually grown since then on the basis of improving assessment in the music classroom.

Various authors have published books on successfully incorporating reliable and valid assessments in secondary music classrooms (Kimpton & Harnisch, 2008; Kimpton & Kimpton, 2013; Kimpton & Kimpton, 2014; Odegaard, 2009). Within all published practitioner’s texts, suggestions and curriculum/assessment creations centered around establishing goals based on national, state, or district/local standards. Additionally, each text, whether explicitly stated or implied, places heavy emphasis on teachers being assessment literate, as previously defined in this paper. For example, Kimpton & Kimpton (2014) suggests forms of writing assessments for secondary music courses

compliant with national common core standards, such as sequencing the rehearsal process, comparing/contrasting performances or pieces of music, and descriptive writing about performances or pieces of music. (pp. 72-74). Additionally, each of these texts contain appendices complete with example curriculum sequences and timelines, graphic organizers for writing assessments, and multiple-choice assessments.

Of particular note, Kimpton, P., & Kimpton, A. K. (2013) address the philosophical purpose of grading and encourages educators to reflect on their grading practices by charting their student assignments in either one of two columns (academic or nonacademic) and determine the weight of each in the student's overall grade (p. 35). Presenting secondary music assignments and grades in such a dichotomous manner is at the center of the present research. The authors promote teaching mastery over performance, drawing from Bernard Weiner's "attribution theory," which segregates learners into opposing groups: the desired mastery-oriented, who welcome challenging learning experiences, and the performance-oriented, those who strive to merely appear competent in comparison to others. These publications (Kimpton & Harnisch, 2008; Kimpton & Kimpton, 2013; Kimpton & Kimpton, 2014; Odegaard, 2009) all suggest non-achievement based assessment is deeply rooted in the culture of music education and strongly recommend a paradigm shift by implementing various practices seen in other subject areas.

Similar to the previously mentioned practitioners' texts, Pellegrino, Conway, & Russell (2015) explain other assessment tools, such as rating scales, checklists, rubrics, and their various combinations. The authors lead music teachers through various steps in creating these assessments, advising educators to break down desired tasks into

observable constructs and placing numerical value to the performance of each construct (pp. 50-51). Regardless of preferred assessment tool, teachers should prioritize the validity and reliability of their practices. Furthermore, the article promotes consistency and authenticity of assessment tools and their administration, uniformity of practice with departmental colleagues, and transparency in planning and responsiveness in reporting results to students, parents/guardians, and administration (Pellegrino et al., 2015, p. 52). All are valuable guidelines and preferred practices from both an administrative perspective and in the view of the academic measurement community.

Perceptions of Assessment in Music Education

Practitioners' texts regarding assessment practices are prevalent in music education literature. In-service teachers are likely to have taken college courses and received professional development trainings on these preferred practices and their supporting philosophies at various times in their careers. However, familiarization with recommended methods in the literature and attendance of courses and trainings do not directly translate to actual practice in music classrooms. Nor do these factors influence the perception of assessment in music education by music students, parents/guardians of music students, and school administrators. The following section of this study will explore how all stakeholders perceive assessment in the music classroom.

McCoy (1991) sought to compare the grading practices of secondary music teachers (band and chorus) to the grading systems preferred by school principals at randomly selected high schools in Illinois. The questionnaire used in this study distinguished between four different grading criteria and included Benjamin Bloom's three domains of learning placed. The three domains—cognitive, psychomotor, and

affective— where placed in a music education context. Cognitive pertained to the factual knowledge of music, psychomotor to musical performance tasks and skills, and affective implied student attitude and participation throughout lesson (Bloom, et al, 1956; McCoy, 1991). The instrument also references non-music criteria, which is described as student attendance and behavior in the classroom. Subjects were provided a list of 25 criteria for grading, teachers and administrators were then asked to indicate the criteria they used or believed appropriate for use respectively. Additionally, subjects were asked to provide a percentage for each criteria to indicate its effect on a student's overall grade. Results show that principals place most weight (40.87% of student grade) on psychomotor criteria, whereas as music teachers reported placing most weight (42.84% of student grade by band directors) on non-music criteria (p. 185). This study shows a significant disconnect between administrators who expect grading practices to reflect student academic achievement and the grading practices actually used by music teachers who seemed to used grades as a form of classroom management. Considering this article appears early in the literature, it is likely these music teachers had limited if any college courses or trainings on recommended assessment practices in the music classroom.

Similar to school administrators, music students and parents/guardians also prefer grading that reflects curricular achievement and mastery course objectives. Conway & Jeffers (2004) examined the perceptions of assessment practices by students, parents/guardians, and the music teacher in three elementary instrumental music classes over the course of approximately two years. Throughout the class students and parents/guardians were provided a detailed music education report card, which outlined student achievement via rating scales and checklists in categories such as duple/triple

meter, executive skills, expression, tone and rhythm (pp. 17-18). In the spring of each year parents/guardians were asked to complete an evaluation of the music education report card used for the class. Thereafter interviews were conducted with parents/guardians who expressed dissatisfaction with the report card. The parents/guardians interviewed generally stated that though they were pleased with the amount of detail in the music education report card, they struggled to understand the achievement of their student in relation to an expected norm or in comparison to other students in the class. Additionally, five students were select for a focus group interview regarding their perceptions of the grading system used in their music classes. The use of the detailed grading system proved to be mutually preferred by students and teachers. Students provided positive responses, feeling it allowed for easier self-assessment. The music teacher noted the system allowed him to avoid grading students on non-music/non-achievement criteria such as “students’ attendance and attitude and other peripheral items” (p. 20). This suggests music teachers may not prefer the subjective assessment commonly associated with their classes but defer to these unreliable practices due to limited understand of recommended alternatives.

Zhang & Burry-Stock (2003) explores the assessment practices of teachers in various levels and subject areas as well as teachers’ perceived preparedness for assessing students. In completing the Assessment Practice Inventory, teachers were asked to report the frequency of use and skill in administering 67 assessment practices. The survey responses indicated assessment practices vary by level and content area. Secondary teachers reported higher use of objective assessment practices, such as paper-pencil tests, and more concern with assessment validity and reliability as opposed to primary school

teachers (p. 335). Secondary math and science teachers used non-achievement grading more frequently than other academic areas (p. 332). Most saliently, educators with more training in assessment and measurement have higher self-perceived skill in assessment and grading, regardless of years of experience teaching (p. 335). This final finding has been observed in other studies, however further research would determine the effect of continued assessment education and professional development on reliable assessment practices in the classroom (Kotora, 2005; Pierre & Wuttke, 2015; Zhang & Burry-Stock, 2003). The use of subjective or less reliable non-achievement student data in assessment and subsequently in determining students' grades is not exclusive to elective classes, though teachers of elective courses may use these discouraged practices more frequently. The following section will address the use and consequence of relying on non-achievement criteria.

Grading Using Non-Achievement Criteria

It is well documented that music education practitioner texts, school administrators, and the measurement community at large strongly advise against using non-achievement criteria for assessing students. These practices can be problematic due to their inherent subjectivity and obvious threats to validity. Literature engages the discussion over grading students on participation at the collegiate level as well as the resistance to and challenges of grading on student achievement in performance-based courses, such as music and physical education.

Early in the literature on the use of non-achievement criteria, Cross & Frary (1999) addressed the practice in a larger categorization of grading referred to as "hodgepodge grading," wherein teachers weigh students' attitude, effort, and

achievement to determine grades. The researchers cite a variety of reasons why teachers might employ such a grading system, including grade inflation (to pad for low achievement, but high effort), or to avoid a high number of failing grades, both of which benefit students. Despite any student-centered reason for hodgepodge grading, the authors and measurement community at large strongly recommend “technical purity” in grading so as to communicate reliable data of student achievement (p. 55). Via a survey of 152 general education middle and high school teachers (specials or “non-academic” responses were withdrawn from analysis), Cross & Frary (1999) found that despite the belief report cards should report student achievement, effort, and attitude separately, teachers indicated their grades reflect either student effort (66% of teachers), or student conduct or attitude (39% of teachers) (p. 63). This study finds all teachers, despite their content, experience, or assessment-based training, either incorporate or at least consider student non-achievement criteria when determining student grades. Though later research would dispute this assertion based on in-service teachers’ net time spent in assessment courses or professional development, the observable use of non-achievement in all classes confirms the pervasiveness of this ill-advised assessment practice.

Similarly, Randall and Engelhard (2010) sought to investigate how secondary school teachers determine grades when considering student ability, achievement, behavior, and effort. A sample of 516 public school teachers were provided a survey with multiple grading scenarios where hypothetical student displayed various levels of ability, achievement, behavior, and effort. For each item, the hypothetical student’s grade was on the borderline between two letter grades (A or B, B or C, etc.) and teachers were required to select either the higher or lower grade based on the details of the scenario (p. 1374).

The results of the study show that teachers were less likely to reward extra percentage points to students who demonstrate low effort and poor behavior. Additionally, results indicated teachers were more likely to grant well-behaved but low-achieving students who demonstrate high effort a higher borderline grade. Lastly, Randall and Engelhard (2010) found that students with average or excellent behavior were more likely to receive a passing grade from teachers despite demonstrating low effort, ability, and achievement (p. 1376). The results of this study confirm the use of non-achievement criteria for grading in general education classes, despite recommendations stating student grades should only reflect achievement measures. Particularly, student behavior was shown to be the most common and influential grading factor in cases where students were on the borderline between two grades.

The reliance on non-achievement student data is overwhelmingly found in the subject area of physical education, a content where students are expected to demonstrate understanding by achieving a performance objective similar to music education (Johnson, 2008). In-service physical education teachers were observed basing their grades on “managerial” criteria or “administrative task,” such as student attendance, participation, and effort, despite the recommendations of experts and academics in their field (p. 46). Johnson additionally highlights the use of “pseudo-accountability” measures, where physical educators are inclined to grade on a performance task, but a majority of the grade’s weight is placed on effort and behavior, rather than performance task achievement. To curb these subjective practices, the literature advises physical education teachers to focus assessment on student mastery and to approach the student’s grade with positive scoring, where students earn their grade by achieving an objective.

The use of non-achievement grading is neither exclusive to any specific academic content area or level of education. The same assessment practices already described are debated at the university level, where grading students on participation is common. Mello (2010) presents positive and negative consequences and implications regarding the use of class participation as an assessment. This review will focus primarily on the identified faults of this practice. Beyond the common criticism regarding the subjectivity of grading participation, Mello cites various additional problems inherent to such grading practices. A notable criticism is the absence of any tangible evidence of students' participation after the fact, which complicates a teacher's ability to defend their grades against any disputes from a student. The second criticism focuses on how grading participation rewards performance rather than learning. The findings reflect the recommended use of performance objectives found in both music or physical education. However, the lack of their use in most university level discussions sections leaves students unaware of achievement standards and essentially requires instructors to rely on non-achievement data for determining student grades. In this practice, hypothetically speaking, a well-prepared student with difficulty communicating or performing in a class environment would be disadvantaged, despite have achieved the learning objective. Mello states this same criticism can be applied for students from cultural backgrounds which may promote "passive compliance" in the educational environment (p. 79). The final criticism of assessing class participation involves a common lack of communication from instructors regarding the grading parameters. Mello claims that students are often "unaware of what constitutes acceptable participation for the instructor" (p. 80). Though the concerns addressed in this article focus exclusively on higher education courses, the same

reasoning can be stated against the use of non-achievement criteria in secondary education classrooms.

Both the insights and concerns regarding non-achievement grading practices in general education classes, physical education, and higher education can be seamlessly applied to secondary music education. In performance ensemble classrooms, music teachers' use of non-achievement grading criteria, such as concert attendance or undefined class participation, likely disadvantages students who lack the means to transport themselves to performance venues or students with cultural norms which run counter to expectations of the daily rehearsal setting (Bradley, 2015; Fautley, 2015; Hoffman, 2013). Though these implications are found in the literature, various studies find in-service music teachers continue to use unreliable subjective grading practices in their programs despite experience or classroom, school, or district teaching context.

Assessment Practices of In-Service Music Teachers

Over the past three decades, researchers have collected data on the assessment practices of in-service music teachers in both primary and secondary schools. The results of multiple projects, each with different empirical intents and methods, point to a wide range of grading practices being used in American music classrooms. This study is particularly interested in the practices of secondary music teachers, specifically those teaching performance ensemble choruses; the following literature will focus on similar studies.

McCoy (1988) is the earliest scholarship specific to the field of music education and sought to collect data on the types of grading being used in high school band and chorus classes in Ohio. The study also sought to categorize assessment practices into

Bloom's three domains of learning (cognitive, affective, psychomotor) and non-music criteria, finding that most teachers' grading reflect a combination of these categorizes. However, the most commonly used indicator of student grades was the non-music category, which includes grading students on attendance, behavior, and instrument care (p. 16). Teachers' grading practices were not influenced by ensemble size, selectivity, or frequency of performances, implying teacher preference as the driving effect on assessment. McCoy (1988) additionally found that teachers who incorporated cognitive domain assessments were more likely to provide students and parents/guardians with course objectives and explanations of their grading procedures. Band and chorus teachers who received grading guidelines from their administration incorporated more assessments based on a psychomotor domain and used fewer non-music criteria for determining student grades.

Kotora (2005) sought to survey the use of different assessments both practiced by in-service high school chorus teachers and taught by college choral methods professors in the state of Ohio. This study provided teachers and professors with a questionnaire containing twelve different forms of assessment, including but not limited to video recordings, audio recordings, written tests, student portfolios, rubrics, concert performances, and student attendance. The most widely used forms of assessment by responding high school teachers included concert performances, student participation, and student attendance (p. 71). In contrast, the most commonly taught forms of assessments by undergraduate choral methods professors were video recording, written tests, concert performances, and student attendance (p. 71). In both cases, a majority of respondents indicated the reason for using these practices was personal choice.

Kotora (2005) asked participating high school chorus teachers to indicate the extent to which their undergraduate choral methods course prepared them for conducting assessments, whereas collegiate choral methods professors were asked to rate how well they believed their course prepared future teachers for assessment (Kotora, 2005). This data was collected using a 5-point Likert scale, which arranged (from lowest to highest) preparation for assessment as “not at all,” “not much,” “not sure,” “somewhat,” and “very well” (p. 73). Most high school teachers (41%) felt their undergraduate methods prepared them “not much” for assessing students; a full quarter of participants indicated their undergraduate methods prepared them “not at all.” Conversely, 55% of college choral methods professors responded that their courses “somewhat” prepared their students for assessment, while 30% indicated their courses prepared future teachers “very well” for assessment in the chorus classroom. Though these findings dispute claims made by other scholars regarding assessment-literacy (where amount of assessment training and guidance increases a teacher’s confidence in assessment and use of reliable grading practices), Kotora (2005) confirms a need for post-undergraduate instruction and continued professional development and assessment training for in-service music teachers.

Questioning teachers about their knowledge of preferred methods for assessing students in music courses constitutes another research approach. In this vein, Pierre & Wuttke (2015) questioned in-service music teachers on their awareness and use of Standard-based grading (SBG), a criterion-referenced assessment practice where student success is determined only by their achievement of specific objectives relating to the content. This grading practice is similar to those advocated by physical education

scholars looking to improve assessment in their field (Johnson, 2008). Participating music teachers were required to complete a survey with questions regarding demographics, knowledge/understanding of SBG and rationales for using or not using SBG in the music classroom. The survey items featured both open-ended response and “yes/no” checkbox formats. Results of this study indicated that a small majority (52%) of teachers were unfamiliar with SBG. Interestingly, of those claiming to be familiar with the grading practice, only 73% were able to adequately communicate their understanding of SBG in an open-ended response. Considering this data, Pierre & Wuttke infer approximately 40% of all respondents adequately use standard-based grading practices in their music classrooms (pp. 5-6). SBG was most commonly used by teachers, the most common main reason for use was due to a set requirement. Separately, those not using SBG indicated a lack of knowledge regarding the practices as the main prohibitive factor for not incorporating the grading method in their classroom (p. 7).

In review of the literature on assessment practices of music teachers, Russell, J. A., & Austin, J. R. (2010) is comparatively the most comprehensive recent study regarding the topic. This study sought to both survey district-level and classroom-level structures for secondary music education and the assessment practices of secondary music teachers, while also investigating any contextual variables that influences the types of assessment used (p. 40). The researchers employed an instrument which questioned teachers’ school district policies regarding grading, assessment practices used and their respective weights in students’ overall grade, classroom/program structure and teaching background. Respondents’ results to the first section of the instrument indicated that 95% of teachers worked in districts where students receive traditional letter grades and 83% of

teacher's course grades were equally weighted to figure student grade point average (p. 42).

Regarding to assessment practices used, Russell & Austin found most music teachers grade students on performance, attitude, and attendance, which confirmed a majority of secondary music teachers continue to grade student on non-achievement criteria. Results showed the average grade weight for achievement criteria like performance/skill was 28% of the students' grade, whereas non-achievement criteria, such as student attendance, was 25%. The biggest indicator of the attendance grade was the student presence at school concerts, with the average being 10 large performances a year. For achievement assessments, most teachers (97%) used written practices such as quizzes and worksheets to determine grades for student's musical knowledge of notation and terminology. A high majority of music teachers (82%) used performance exams to determine grades for technique/skill, with some using tools such as rubrics and rating scales while others used informal "global" observations (p. 45).

By analyzing correlations between assessment practices used and school/classroom context, Russell & Austin found "several significant correlations with weak to modest magnitude" (p. 47). Specifically, music teachers with administrative guidance over grading were less likely to place heavier weight on assessments of student attitude. This correlation was also found among teachers with more weekly instructional time. Music teachers who had more performances throughout the academic school year were less likely to prioritize grading students' musical knowledge, rather focusing more on non-achievement criteria, such as student attendance. Lastly, teachers who expressed high confidence in conducting student assessments were more likely to assess student

performance (p. 48). These results present music teachers grading practices in a similar situation to those used by general education teachers, where both achievement and non-achievement criteria are used to determine student grades. Music teachers, however, are shown to place heavier overall grade weight on non-achievement criteria than their general education counterparts. This comparison, paired with findings regarding the disadvantages of using unreliable and subjective grading practices, justifies the need for further research in the area of assessment practices in secondary music education in an effort to remediate the frequency of these ill-advised methods.

Due the comprehensiveness of Russell & Austin (2010) and the lack of replication in the social science literature, the present study will be a replication of Russell and Austin's project. Though replication studies are often conducted in the hard sciences, where their common use is expected to confirm findings and legitimize studies, social sciences replication studies appear significantly less frequently. Mackey (2012) concedes that results from social science projects can be difficult to replicate, but argues nonetheless that increasing the number of replications will improve overall scholarship within the larger field (p. 21). The problem lies in how social science researchers work to replicate previous studies, as various forms of replication exist and are invoked for methodical reasons. Schmidt (2009) distinguishes between two separate approaches to replication research: direct replication, which consists of an exact recreation of the previous study procedure, and conceptual replication, those involving near-recreation of the study with minor changes or where a researcher attempts to reproduce the result of the previous study through different methods (p. 91). Mackey (2012) suggests direct or exact replications are nearly impossible to conduct in the social sciences, as human

samples and variables or stimuli are difficult to precisely replicate. Despite the limitations, replication is still needed in the field of social sciences, specifically in the area of music education. Russell & Austin (2010) provide a reliable and valid measurement tool and a clear methodology; both are highly conducive to a conceptual or approximate replication of their work.

Chapter 3

Method

The purpose of this study is to examine assessment practices of secondary music teachers utilize in their performance-based ensemble classes. The review of the literature revealed a need for continued examinations of music assessment practices as well as a gap in the music education scholarship regarding replication. These vacancies led me to replicate the Austin & Russell (2010), by employing their instrument and research sequence. In addition, the guiding research questions for this project are identical to those from Austin & Russell (2010) in order to extend the research to a different geographic region and achieve comparable results to the previous study. The research questions are below:

1. What types of school district frameworks and classroom contexts are secondary music teachers operating within as they assess learning and grade students?
2. Which specific assessment and grading practices are most commonly employed by secondary music teachers?
3. Do any contextual or individual difference variables influence secondary music teachers' assessment and grading practices?

Replication Guidelines

The framework of replication in social science proposed in Schmidt (2009) are used as guidelines to ensure the validity of this study's replication status. Austin & Russell (2010) provides clearly delineated methods complete with a reliable instrument for this project to pursue what Schmidt (2009) defines as a "direct replication." In that, the specific function of this study is to generalize the previous results to a different

population in an effort to expand the knowledge on this topic (Schmidt, 2009, pp. 93-95). Regarding this format of replication, Schmidt (2009) states the primary information focus, contextual background, and constitution of dependent variables all remain consistent between both studies. To follow these guidelines, the present study will shift the geographic focus to the Mid-Atlantic region of the United States.

Any other changes between the initial study and this project are seen as inconsequential, such as the digitization of the hard copy instrument and the method for distributing the instrument. These changes were perceived as necessary to improve participant completion time and response rate. As these modifications are not expected to create any threats to validity or fundamentally change the original focus or methods of the project, thus they are permissible within the framework for direct replication as provided by Schmidt (2009).

Instrument

As this study is a functional approach to replication of Russell & Austin (2010), the same instrument was used to collect data regarding the assessment practices of secondary music teachers. This instrument, the Secondary School Music Assessment Questionnaire (SSMAQ), was originally designed and implemented by Austin (2003), after which it was adapted for Russell & Austin (2010). The original questionnaire was built in the image of various instruments used in similar studies focused on the assessment practices of teachers, such as Cross & Frary (1999) among others (Austin & Russell, 2010, p. 41). To improve validity, the instrument's items were designed to more specifically pertain to the content of music education, as other questionnaires' items were worded for use of general education teachers. The first draft of the adapted SSMAQ was

subject to a trial run by 10 veteran music educators. This pilot resulted in Austin & Russell (2010) removing 10 items due to redundancy or insignificant data return, additionally remaining items were reworked to decrease indirect wording and response time (p. 41). The finalized version of the SSMAQ relies predominantly on check-lists and multiple-choice format, with minimal short-answer items. Austin & Russell (2010) anticipated the average completion for the questionnaire would be 8 minutes, as participants would be able to skip multiple items that do not pertain to their grading practices (p. 41).

The SSMAQ contains three sections, two pertain to specific assessment policy and practices, whereas the last inquires about music classroom demographics and specific teaching contexts. Section one focuses on secondary teachers' grading frameworks, particularly district grading structures for their and how performing ensemble course grades affect students' overall grade point averages (GPA) and graduation requirements. The only text-based question of this section requires teachers to indicate the percentage of students who receive each letter grade. Additionally, few items require participating teachers to indicate administrative guidance and the influence of standards-based curricula using a Likert scale. The second section, titled "Specific Assessment Strategies," first asks teachers to share approximate weights they give to various criteria for determining students' overall grades in their performance ensemble classes. Thereafter the questionnaire is broken into five subsections containing questions about each criterion used to determine students' overall grade (i.e. "Attendance and Punctuality," "Attitude," "Written Assignments," "Practice Assessments," and "Performance Assessments.") Participants are encouraged to only complete sections that

pertain to their grading practices. I modified these sections of the SSMAQ by adding an “other” option for various items with participations given the ability to provide explanation in a text field.

The third and final section of the SSMAQ contains items regarding teaching context, such as type of performing ensemble class taught, teaching level, and various metrics regarding the teachers’ music program. Additionally, this sections asks participants to indicate their assessment confidence on a Likert scale, their music teaching experience measure in years, highest degree earned, and professional development experience. The questionnaire concludes by asking if participants are interested in answering additional questions regarding their assessment and grading experiences.

Though I had no intention of editing or modifying the content of the SSMAQ beyond the addition of the “other” option to various items, a digital version of the questionnaire was constructed with the web-based survey tool, Qualtrics. Access to this platform was provided by the University of Maryland, College Park. This project’s digital SSMAQ utilizes skip logic which directs respondents to specific items based upon answers to previous questions. Allowing the questionnaire to be accessed and submitted digitally was prioritized to make distribution of the instrument and responding for participants more convenient. With that said, this project maintains the Russell & Austin (2010) estimate of an 8-minute completion time for the SSMAQ.

Participants

The target population for this study was secondary music teachers whose primary teaching assignment is a performance-based ensemble course, preferably band, chorus, or

orchestra. However, this project's geographic focus was the Mid-Atlantic region of the United States, which includes Delaware, Maryland, Pennsylvania, Virginia, and West Virginia. This sample was generated from a list of Mid-Atlantic NAfME members who teach secondary band, orchestra, and/or chorus during the 2016-2017 school year. A total of 4,083 invitation emails were sent via NAfME's email transmission platform once per week over the course of three weeks. The first distribution produced 136 responses, the second distribution added 94 responses, and the final distribution added 61 responses before the completion window closed. A total of 291 responses were received, however, of this total, only 188 respondents completed the instrument in its entirety. Thereafter, three duplicate responses were found during the data cleaning procedure. These duplicate entries were removed from the dataset at random through a listwise deletion process. Subsequently, the project was left with 185 usable responses.

It is difficult to determine this study's response rate due to a variety of factors. NAfME's email transmission platform sent a total of 4,083 invitation emails, however the organization's distribution statistics reported only 840 emails were opened. Additionally, I discovered invitation emails were distributed to individuals out of the intended population. Various out-of-scope colleagues and university professors confirmed receiving my invitation emails through NAfME's service. These factors led me to conclude this study's true response rate is incalculable. In place, I propose a response rate range based on the total invitations sent and the emails reported opened. Thus the lower bound response rate is assumed to be 5% and the upper bound is placed at around 22%.

The majority of teachers in the sample (51%) reported band as their primary teaching assignment, 39% teach chorus, and 10% teach orchestra. Regarding teaching

level, 51% of respondents teach in high schools and the remaining 49% of respondents teach in middle schools junior high schools. The average teaching experience of responding teachers was between 10 – 20 years in the profession ($M = 16.87, SD = 10.6$) with total teaching experience of the sample ranging from 0 – 43 years. The majority of participating teachers (63%) indicated having master’s degrees, 31% held bachelor’s degrees, and remaining respondents (7%) achieved doctorate degrees. Responding teachers reported to receive additional assessment training through conference clinics (83%), district in-service trainings (64%), graduate courses (61%), and university workshops (37%). The SSMAQ comments section revealed few teachers gained supplemental assessment education through the Nation Board Certification process ($n = 2$) or individual research through academic journals and assessment readings ($n = 3$).

Table 3.1

Distribution of Sample by State Compared to Distribution of Mid-Atlantic Active NAfME Members as of January 2017

State	Sample	Active Members
Delaware	3%	3%
Maryland	6%	6%
Pennsylvania	15%	12%
Virginia	34%	32%
West Virginia	41%	47%

When segregated by state, Pennsylvania and Virginia produced the highest number of respondents with 41% and 35% of the sample reported teaching in these states, respectively. The remaining three Mid-Atlantic states produced comparatively small representation in this sample with 15% of participants teaching in Maryland, 6% in West Virginia, and 3% in Delaware. These percentages of respondents by state closely matched

the state breakdown of active NAFME members as of January 2017 in the US Mid-Atlantic (see Table 3.1). Information provided by NAFME shows Pennsylvania and Virginia having the largest active member populations, and West Virginia and Delaware having the fewest. Considering these state-by-state comparisons of NAFME members, this sample is comfortably proportional to the larger population.

Procedure

After receiving approval through the Institutional Review Board, participants were sent the digital versions of the questionnaire via e-mail provided by NAFME membership services. Each invitation e-mail contained an explanation regarding the focus of the study, a guarantee of anonymity, and the deadline for response. The response gathering window remained open for three weeks, with reminder e-mails sent each week in an effort to improve the response rate. Each reminder included the original invitation text and a link to the digital questionnaire. After the response gathering process, raw data was exported as a .csv file and then converted into a Microsoft Excel file for cleaning. Open response answers and the qualitative data were coded also using Microsoft Excel. After all response data was cleaned, the data set was imported into Stata/IC, version 14 for statistical analysis.

This study incorporates both descriptive and inferential statistical models to answer the guiding research questions. Specifically, descriptive analyses were used to answer research questions one and two, whereas inferential statistical measures were used to answer research question three. These were the same models used by Austin & Russell (2010).

The models of descriptive and inferential statistical analyses used for this study include central tendency reporting, pairwise correlation, and multivariate analysis of variance (MANOVA). Most results from the SSMAQ will be used for descriptive central tendency reporting. However, pairwise correlation and multivariate analysis was used to infer significant relationships between teacher grade weigh differences (pertaining to attendance and punctuality, attitude, written demonstration of knowledge, practice measurement, and performance achievement) and contextual/demographic metrics such as school administration guidance, effects of standard-based grading, instructional time, number of ensemble students, number of ensemble performances, teacher assessment confidence, teaching experience, and highest degree earned. Lastly, due to the small number of usable responses, this project used the alpha level of .05 for all statistical analyses.

Chapter 4

Results

The following chapter will present results gathered by the Secondary School Music Assessment Questionnaire from a sample of Mid-Atlantic secondary school music teachers. This data is organized in similar fashion to those reported in Russell and Austin (2010), so to allow for parallel comparison of the two projects. As mentioned in the methods chapter, both descriptive and inferential statistics will be reported in this chapter. Descriptive results regarding district-level policy are presented first, followed by school-level grading structures, and finally, classroom-level assessment practices. This organization illustrates the top-down influences of district assessment policies on classroom teaching practices. Additionally, descriptive results will be disaggregated by state, teaching assignment and level to a further extent than the results reported in Russell and Austin (2010). Inferential statistics, including central tendency reporting, pairwise correlation, multivariate analysis of variance (MANOVA), and analyses of variance (ANOVAs) are presented in the final portions of the chapter.

Descriptive Statistics

District-level results. SSMAQ items regarding district-level policy and frameworks inquired about the use of traditional letter grades and percentages, as opposed to standard-based grading, pass/fail structures, or no grading systems at all. The majority of respondents (95%) indicated their students are graded using traditional letter grades and percentages. Alternative district-level grading systems such as standard-based grading and pass/fail formats were reported by 3% and 2% of all respondents, respectively and only one respondent indicated having no designated grading system.

When disaggregated by state, all respondents from Delaware and Maryland reported the use of traditional letter grades, whereas Pennsylvania, Virginia, and West Virginia had few respondents indicating the use of pass/fail or standards-based formats. Three Pennsylvania and two West Virginia respondents claimed using pass/fail grading. Standards-based grading was reported by two Virginia and one Pennsylvania teachers. Overall, these responses are similar to results from Russell and Austin (2010); traditional letter grade reporting is pervasive amongst these respondents in the Mid-Atlantic region and the Southwestern region of the United States.

A separate SSMAQ item inquired about the effect of standards-based curricula on assessment practices using a Likert scale. 47% respondents indicated standards-based curricula having somewhat of an effect on their teaching practices, 25% indicated no effect at all, and 17% indicated quite a bit of an effect. Overall, these descriptive statistics are similar to the results reported in the original project, as 71% of all respondents are claiming standards-based curricula having little to no effect on assessment. Lastly, 9% of all respondents stated standards-based curricula has not been implemented in their districts. When desegregated by state, 23% respondents from Virginia indicated standards-based curricula having quite a bit to extensively of an effect on their grading practices.

A majority of respondents (68%) stated their music ensemble grades are weighted equally with other general education classes when calculating students' overall grade point average (GPA). Another 24% of involved music teachers reported their grades effect students' GPAs, but are not weighted equally. The remaining 9% of respondents' secondary music course grades have no effect on their students' GPAs. In addition, the

majority of participants (78%) teach in districts where secondary music ensemble courses provide credit towards fulfillment of graduation requirements. This percentage vastly increases to 95% when disaggregated responses by teaching level and focusing on high school teachers, however a majority of middle school/junior high music ensemble teachers (62%) also stated their courses provide credit for graduation. In comparing these district-level results to those of Russell and Austin (2010), consistent similarities in policy are found.

School-level results. School-level items focus on school administrators' monitoring or guidance of assessment practices. Participants were asked to indicate on a Likert scale the extent to which their administration monitor or guide their assessment. A majority of respondents (57%) claimed their administration somewhat monitored/guided their assessment practices, whereas another 35% of participants reported no administrative monitoring/guidance regarding assessment. This reporting on the lack of administrative monitoring/guidance in assessment practices is consistent across teaching level and state.

Classroom-level results. The next area of focus is the classroom-level, with participants reporting specifics about their assessment practices, teaching contexts, and classroom demographics. The average amount of instructional time among all respondents was 176 minutes per week, however instructional time ranged from 30 to 1,000 minutes per week. Regarding number of students, on average, participants reported teaching and assessing 116 students, with a range from 1 – 600 students. Lastly, regarding teaching contexts, participating teachers indicated having to prepare for an average of 8 major performances per year (response range = 2 – 40).

In addition to providing data on teaching context, participating teachers submitted data regarding their grading practices, including assessment objectives and formats, grade weights and grade distributions. The first question of the SSMAQ inquired about communicating course grading policy to students. The vast majority of respondents (87%) reported providing students with their grading policies in written format and another 7% communicate it verbally. However, 5% of participating teachers do not communicate their grading policy to students.

The SSMAQ allowed respondents to indicate their assessment criteria, and the weights they assign to each criteria. The categorizations for assessment criteria include non-achievement measures, such as student attendance, attitude, and practice documentation. Separately, student performance ability and written demonstration of knowledge are categorized as achievement criteria. As found in Russell and Austin (2010), respondents to the project use a variety of assessment criteria defined as “hodgepodge” grading by Cross and Frary (1999). Performance criteria was the most frequently used criteria, in addition to receiving the most weight on average. Attitude was recorded as being the second most common criteria. On average, respondents’ graded practices reflected near equal parts achievement and non-achievement criteria. However, the grade weight for both performance and attended criteria ranged from 0 - 100% of students’ grades.

Participating teachers who used traditional letter grades ($n = 176$) were asked to share their grade distributions for their students. Similar to the findings of the previous study, on average, a high majority of the secondary music ensemble students reflected in

this project receive a grade of A (81%). Thereafter, an average of 12% receive Bs, and only 7% receive Cs or lower.

The SSMAQ returned data regarding specific assessed objectives and assessment formats for the various graded criteria. Assessment categories identified as non-achievement measurements, such as attendance, attitude, and practice, are addressed first. Among teachers using attendance and punctuality metrics for grading ($n = 125$) a high majority (94%) grade students on attending major school performances (see Table 4.1). Grading students on daily attendance and punctuality is also common with 79% and 73% reporting the use of these practices respectively. In addition, 60% of respondents using attendance criteria claim to partially reduce a student’s grade for an unexcused absence and another 36% reduce the grade by one grade or more.

Table 4.1

Assessment Criteria and Average Weights Used in Grading Policies

Criteria	# teachers using	Grade weight		
		<i>M</i>	<i>SD</i>	Range
Achievement				
Performance	176 (96%)	37%	22.93	0 – 100%
Written Knowledge	136 (74%)	12%	12.49	0 – 50%
Non-achievement				
Attendance	125 (68%)	19%	23.29	0 – 100%
Attitude	151 (82%)	23%	19.80	0 – 92%
Practice	89 (48%)	8%	10.34	0 – 50%

Attendance and attitude criteria. Among respondents assessing student attitude ($n=151$), the most commonly used metrics within the criteria were in-class participation,

responsibility (such as bringing required materials to rehearsals and concerts), and effort. Citizenship, associated with courtesy and cooperative behavior, also ranked high as a common factor contributing to students' attitude grade (see Table 4.2). Regarding methodical approach to attitudinal assessment, the majority of respondents (79%) indicated using a combination of subjective and objective impressions as a means of measurement.

Table 4.2

Factors Considered by Music Teachers When Assessing Attendance (n = 125) and Attitude (n = 151)

Attendance factors	% of teachers	Attitude factors	% of teachers
Major school performance	94	In-class participation	89
Daily rehearsals	79	Responsibility (Bring materials)	85
Punctuality	73	Effort	81
After-school rehearsals	56	Citizenship	70
Solo or large-group festivals	28	Instrument/uniform care	34
Athletic events	22	Leadership	30
Other	10	State festival participation	15
		Honor group participation	12
		Private lesson participation	6
		Ensemble support activities	4
		Other	4

Practice criteria. Student practice records were found to be the least common form of assessment criteria between participating teachers, with 48% of respondents reporting its use. When use of practice criteria was broken down by teaching specialization, virtually no difference between instrumental teachers (49%) and chorus

teachers (47%). Among these respondents ($n = 89$), quantitative paper reports or practice cards documenting student practice time and written qualitative reports describing how and what students practiced were reported by 43% and 42% respectively. A quarter of teachers who incorporate practice criteria in their grading indicated using recordings of students' practice sessions as an assessment practice.

Table 4.3

Percentage of Teachers Using Specific Objectives and Formats for Written Assessment ($n = 136$)

Objectives	% of teachers	Formats	% of teachers
Music terminology	89	Quizzes	77
Analyzing/evaluating music	79	Worksheets	68
Identify music elements	60	Journals	39
Music theory knowledge	56	Homework assignments	33
Performance/pedagogy knowledge	45	Projects/presentations	32
Cultural context knowledge	43	Exams	25
Music history knowledge	38	Other	12
Small scale compositions	19		
Compositional technique compositions	16		
Other	3		

Written criteria. Regarding assessment criteria categorized as achievement-based, written assessment practices were used by nearly three quarters ($n = 136$) of all respondents. The majority of these teachers (89%) assessed students on knowledge of music terminology, symbols, and notation (see Table 4.3). 79% of these respondents assess students on the written demonstration of the ability to analyze or evaluate musical

performances. The other most commonly reported objectives for written assessment were ability to identify musical elements and knowledge of music theory principles, such as intervals, chords, voice leading and chord progressions. Quizzes and worksheets were identified as the preferred formats of written assessment with high majorities of these respondents (77% and 68% respectively) reporting their use. Alternative written assessments to worksheets include journals (used by 39% of respondents), homework assignments (33%), and projects or presentations (32%). Only a quarter of these respondents indicated using exams as a format for written assessment.

Table 4.4

Percentage of Teachers Using Specific Objectives and Formats for Performance Assessments (n = 176)

Objectives	% of teachers	Formats	% of teachers
Technique (scales, etudes)	69	Playing exam, live, in class	80
Prepared performance of ensemble music excerpts	72	Ensemble concert performance	66
Prepared performance of solo/chamber repertoire	53	Sectional performance in class	56
Sight reading	46	Playing exams, audiotaped	34
Memorized performances	20	Playing exams videotaped	29
Improvised performance	3	Playing exam, live, out of class	30
Other	7	Auditions	29
		Large group festival ratings	14
		Chair challenges	7
		Solo/ensemble festival ratings	6
		Other	5

Performance criteria. Performance assessment criteria was found to be the most commonly used grading practice among participating teachers ($n = 176$). Student-prepared performances of ensemble music assessing tone, accuracy, and musicality, was the preferred objective for performance assessment among respondents (see Table 4.4). Second to this, student performance technique, such as scales and etudes, was used by 69% of teachers using performance criteria. The most common format for assessing performance indicated by responding teachers was in-class playing exams. A majority of respondents also reported using in-concert ensemble performances (66%) and in-class sectional performances (56%) as format for assessing performance criteria. Lastly, among these respondents, rubrics surfaced as the favored method for scoring performance assessments. Only 17% of teachers indicated relying on global impressions as a means for scoring student performance assessments.

Participants' Comments. I slightly modified the SSMAQ by including an “other” option for various items which allowed participants to indicate separate assessment practices not represented in the instrument. A small percentage of participants indicated “other” for various questions, however their written contributions were often represented in the list of answers or applied to a separate item altogether. One example of these duplicate or inconsistent responses involved a question regarding grade weights per criterion. Though “attendance” was an answer option, a participant selected “other” and commented “concert attendance.” In another case, a different respondent selected the “other” option and indicated “performance assessments, sight reading assessments, scale assessments” as a practice used to assess attendance. Despite these few incongruent

answers, most qualitative answers were duplicated in the quantitative metrics. Thus only comments reflecting assessment practices not represented in the questionnaire were considered and will be addressed.

The amalgamation of participating teachers' comments revealed two themes regarding secondary music assessment. The first being technology as supplemental to participating teachers' assessment practices. Among these few responses, Google Classroom and SmartMusic were referenced multiple times with responding teachers speaking positively about utilizing these websites as grading methods. Inhibiting factors to reliability assessment surfaced as the second theme of respondents comments. Specially, these commenting participants reference time constraints and number of students or class size as complication to their assessment practices. Commenting teachers claim these factors prohibit or complicate their ability to administer reliable assessment practices. These comments reveal a subset of participating teachers feel forced to rely on non-achievement measures due to uncontrollable influences, despite awareness and preference for more reliable achievement-based assessment practices.

Inferential Analyses

MANOVA and ANOVA. As in Russell & Austin (2010), a multivariate analysis of variance (MANOVA) and subsequent analyses of variance (ANOVA) were conducted to infer the effect of teaching specialization (such as band, chorus, or orchestra), level (middle school or high school), or their intersection on assessment criteria grade weights.

The MANOVA only exposed a significant effect of responding teacher's specialization ($\Lambda = .89, p < .05$) on grade weights (see Table 4.5). Thereafter, ANOVAs for each assessment criteria were only conducted for teaching specification, as teaching

level and the interaction between teaching level and specialization failed to produce significant effects. Among the ANOVAs for teaching specialization, only one significant effect was found for the grade weight of written assessment criteria, ($F = 5.59, p < .01$). Further inspection showed that responding chorus teachers gave greater weight to written criteria ($M = 16.1\%, SD = 12.40$) than band teachers ($M = 10.2\%, SD = 12.35$) or orchestra teachers ($M = 8.9\%, SD = 10.23$).

Table 4.5

Multivariate and Univariate Analyses for Significant Relations Between Teaching Level, Teaching Specialization, and Assessment Criteria (n = 185)

	Teaching level $\Lambda = .97$	Teaching specialization $\Lambda = .89^*$	Teaching level/ specialization $\Lambda = .96$
Attendance	–	1.27	–
Attitude	–	0.01	–
Practice	–	0.82	–
Written	–	5.59**	–
Performance	–	1.81	–

Note: Wilk's Lambda (Λ) values in header, F values in cells.

* $p < .05$, ** $p < .01$

Correlation. In addition to the MANOVA regarding grade weights, teaching level, and teaching assignment, correlational analysis was conducted to find significant relationships between grade criteria weights and teaching contexts and demographics. Specific teaching situations in the analysis included administrative assessment guidance, standards-based curriculum adoption, instructional time, number of ensemble students, number of performances, assessment confidence, years teaching, and highest degree earned. Significant interactions surfaced between various grade criteria weights and

teachers' instructional time, number of ensemble performances per year, and teachers' assessment confidence (see Table 4.6).

Table 4.6

Correlation Between Assessment Criteria Weight and Teaching Contexts

Teaching Contexts	Assessment Criteria				
	Attendance	Attitude	Practice	Written	Performance
Administrative guidance	.011	-.060	-.039	.046	.020
Standards-based curriculum adoption	-.119	-.025	.103	.067	.046
Instructional time	-.161*	-.082	-.032	.150*	.157*
Number of ensemble students	-.076	-.039	.008	.075	.076
Number of performances	.161*	-.091	-.064	.061	-.026
Assessment confidence	-.003	-.212**	-.034	-.043	.208**
Years teaching	.025	-.045	-.005	-.033	.027
Highest Degree Earned	-.015	-.062	-.101	.021	.099

Note: * $p < .05$, ** $p < .01$

The correlation analysis suggests participating secondary music teachers with more instructional time are less likely to have high grade weight for attendance ($r = -.161$), and more likely to have higher grade weights for both written ($r = .150$) and performance ($r = .157$) assessment criteria. A weak correlational relationship was found between number of performance per year and the attendance assessment criteria ($r = .161$), where respondents with a higher number of performances to prepare were most likely to have a high grade weight for student attendance. Two interactions with the strongest correlational relationships, though still relatively weak in magnitude, included teacher's assessment confidence and grade weights for attitude and performance. The analysis suggests, though to a minimal degree of significance, participating music

teachers with higher assessment confidence had lower grade weights for attitude criteria ($r = -.212$), and higher grade weight for performance ($r = .208$) when determining students' overall grade.

Chapter 5

Discussion

The ongoing accountability trend in American education places a high premium on assessment literacy among teachers of all contents (Deluca and Bellara, 2013). This reliable assessment prioritization has been propagated by music education journals and practitioners' texts, with authors encouraging music teachers to adopt assessment practices that reliably represent student achievement of curricular goals and objectives. However, few music education researchers have published examinations of music teacher's assessment practices over the past two decades. These studies unearthed revealing inconsistencies of assessment practices between music teachers, as well as a widespread use of grading methods discouraged by music education academics and the measurement community (Kotora, 2005; McCoy, 1988, 1991; Pierre & Wuttke, 2015; Russell & Austin, 2010). The current project is a continuation of this music assessment scholarship, by investigating the assessment practices and grading procedures of secondary music teachers in the Mid-Atlantic region of the United States. This final chapter unpacks and discusses the results from the Secondary School Music Assessment Questionnaire and the subsequent statistical analyses in context with the three guiding research questions regarding school districts frameworks, secondary music teachers' assessment practices, and individual teaching variables. In addition, this chapter compares results and findings with that of Russell & Austin (2010), the original study upon which this project is based.

Research Question 1

This study's first research question addresses district policies and classroom contexts affecting music teachers in the Mid-Atlantic states. The question seeks to examine and describe external assessment frameworks and internal settings where participating teachers deliver their assessment. Traditional letter grade formats were reported by the majority of participating music teachers, with very few responding teachers working in districts with standards-based grading or pass/fail policies. Though few responding teachers indicated working in standards-based grading frameworks, nearly half of all respondents also stated the implementation of standards-based curricula by their districts having somewhat of an effect on their approach to assessment. These findings show an inconsistency of practice and familiarity with standards-based grading similar to findings of other researchers on this topic, where a majority of music teachers claimed to be unfamiliar or had difficulty defining standards-based grading practices (Pierre & Wuttke, 2015).

The majority of participating teachers indicated their music classes were weighted equally with other courses to calculate student grade point averages (GPA) and that their classes provide credits which apply to student graduation. In conjunction with these findings, the majority of responding music teachers claimed to receive little, if any, guidance from their school administration regarding their assessment practices. Again these policies fall in line with those reported by Russell & Austin (2010) suggesting a commonality between the experience of secondary music teachers in the Southwest and the Mid-Atlantic regions of the United States. Additionally, the lack of administrative oversight of music assessment continues to be a common observation in this scholarship

(McCoy, 1988, 1991; Russell & Austin, 2010). However, isolation when devising classroom assessment is not limited to music teachers. More assessment guidance and communication between general education teachers and administration is a common suggestion for improving grading practices (Cizek, Fitzgerald, & Rachor, 1995; Zhang & Burry-Stock, 2003). With respect to specific classroom contexts, participants' instructional time per week, number of students, and amount of major ensemble performances a year ranged widely indicating more inconsistency within the field of music education.

Research Question 2

Research question two centered on finding common assessment and grading practices among responding secondary music teachers. Considering the average of assigned grade weights per assessment criteria, non-achievement grades (such as attendance, attitude and practice) slightly outweighed achievement-based grades. This near balance of assessment criteria signals positive difference regarding the use of reliable grading practices between respondents to this project and the southwestern sample of teachers represented in Russell & Austin (2010). The majority of responding teachers employ written assignments to assess music objectives, though the average grade weight for this form of assessment ranked low in comparison to other assessment criteria. Knowledge of music terminology (definitions, symbols, and notation), and music analysis/evaluation were the most common objectives assessed in this format. These practices gesture towards higher levels of Bloom's taxonomy or Depth of Knowledge being used by participating music teachers, however student composition and journal use continues to be infrequent (Hanna, 2007). Though responding music teachers assess

student achievement of high level music objectives (music analysis and evaluation) through recommended practices and formats, the average grade weight of these forms of assessments are significant in comparison to non-achievement or performance criteria.

Performance surfaced as the most commonly used assessment criteria, as well as the greatest weighted individual criterion on average, among participating teachers. Consistent with the recommendation by practitioner texts, rubrics were reported to be the most preferred form of assessment for student performance. Responding music teachers most commonly assessed students on techniques and material learned in class. This practice reflects the suggestions made physical education regarding performance grades (Johnson, 2008). The common use of rubrics and written assessments among participating music teachers allows for reliable record keeping of student achievement. Though these observations suggest higher assessment accountability among music teachers, the limitations of this study (i.e. narrow sample size, and self-select or non-response bias) should remain in consideration as 81% of responding teachers indicated being very to extremely confident in assessment. In such case, this study shows that among teachers of high-confidence in assessment, less-reliable grading practices are continuously used.

Even with a high representation of assessment-confident teachers in this sample, the continued use of non-achievement measures when calculating music students' grades suggests unfamiliarity or a disregard for specific recommendations from assessment scholars regarding non-achievement criteria (Asmus, 1999; Johnson, 2008; Pellegrino, Conway, & Russell, 2015). The majority of participating teachers who academically penalize students for unexcused absences further emphasizes this disconnect between

classroom grading procedures and recommended practices (McCoy, 1991; Pellegrino, Conway, & Russell, 2015). In addition to grading on attendance, the assessment of student “attitude,” participation, and preparedness proved to highly common in responding teachers’ grading practices. Despite this, the use of these metrics when determining students’ grades has been discouraged by general and music education assessment scholars alike (Asmus, 1999; Johnson, 2008; Kitora, 2005; McCoy, 1988, 1991; Randall & Engelhard, 2010). The present study’s results suggest the continued push for achievement-based assessment measures has led teachers to more frequently incorporate reliable assessment practices which reflect music performance objectives. However, more instruction for music teachers regarding the unique difference between non-achievement and achievement metrics may limit continued assessment of student attendance, attitude, and participation.

Lastly, the grade distributions of participating music teachers indicate a separate, but related concern: regardless of assessment criteria and practices uses, these music teachers reward a vast majority of their students with As. This calls into question whether the secondary music course grade is a truly indicative measure of their musical achievement as opposed to an indicator of consistent attendance, preparedness, and cooperative behavior.

Research Question 3

The third research question sought to find significant relationship between music teachers’ assessment practices and their individual teaching contexts. This project found that teaching specialization slightly influenced graded weights. Chorus teachers, on average, place greater weight on written assessment criteria. Russell & Austin (2010)

posit the likelihood of chorus teachers incorporating written song or text interpretation assignments due to the nature of their music. This difference may also be the result of logistics as vocal music students can more conveniently demonstrate objectives in written form due to the absence of a physical instrument. Also instrumental classes may place a premium on students mastering the technicalities of their instruments rather than devoting in-class playing time to written assignments.

Beyond teaching level or specialization, various respondents in the present study indicated through comments that specific teaching contexts (such as limited instructional time, number of students, and number of performances) had an effect on their assessment practices. These comments, to a certain extent, were verified through the correlational analyses of teaching contexts and assessment criteria grade weights. Among teachers participating in this project, those with more class time were less likely to place heavier grade weights for attendance criteria and more likely have increased grade weights for performance and written criteria. Considering the time required to administer reliable written and performance assessments (especially at the individual level), it is understandable that responding music teachers with more instructional time feel comfortable attributing a larger percentage of their students' grades to written and performance criteria. However, though there is logic behind the practice of teachers with limited class time assigning heavier grade weight to student attendance, this practice fails to utilize grades for their intended purpose of representing student achievement (Asmus, 1999; Pellegrino, Conway, & Russell, 2015).

There were few findings regarding the influence of teaching context consistent with Russell & Austin (2010), particularly among respondents with higher assessment

confidence placing higher grade weight for performance criteria and less weight for attitude criteria. Teachers' self-perceived assessment confidence and assessment practices have been linked to their assessment literacy (Pierre & Wuttke, 2015; Zhang & Burry-Stock, 2003). Such assessment-proficient teachers would prioritize performance assessment for a performance ensemble class and would place less emphasis, if any, on assessing student attitude or participation. Also in congruence with the Russell & Austin (2010), participating teachers with more performances per year tend to have heavier grade weights for attendance criteria. As previously mentioned, this correlation and practice is likely used for incentive purposes, as attendance at major ensemble performances was the most common factor attributing to respondents' attendance grade. Interestingly, the analyses of this study failed to find a significant correlation between assessment criteria and administrative guidance, as found in Russell & Austin (2010).

Comparison Between Original and Replication Studies

As this project is a replication of Russell & Austin (2010), it is salient to compare results of the two studies beyond context of the three research questions. The most interesting finding was the near identical rankings of common assessment objectives and formats for most assessment criteria between the two samples. The rankings of common attendance factors and performance objectives produced matching lists among responding teachers to this project and those from Russell & Austin (2010). In addition, the top three rankings of commonly used attendance attitudinal factors, written assessment objectives, and performance formats corresponded between the two studies. An interesting difference in common practice appeared for formats to assess written and performance criteria when comparing these results. Assessing student journals and

videotaped playing exams proved to be more common among the present study's sample. These differences in assessment practice may be attributed to the increase of convenient technologies, such as photo-capable cell phones and flash-based laptops, in the classroom since the previous study. Overall, the consistencies in common assessment practices outweigh the few differences between the original study and this replication. It appears that Mid-Atlantic secondary music teachers who participated in this project prioritize and most commonly utilize the same set of assessment practices to conduct their performing ensemble classes as those from the Southwestern region of the United States reflected in Russell & Austin (2010). This is difficult to parse, as, despite using similar grading practices, the significant difference exists between the specific weight music teachers assign to each assessment criteria.

Implications and Conclusion

Through this study, I sought to advance the scholarship on secondary music education assessment practices by directly replicating the work of previous researchers in a separate geographical location. Though the project received a considerable number of responses spanning the Mid-Atlantic region of the United States, the small response rate greatly limited the ability to make robust claims about music assessment practices for the population. Despite this, the collected data produced significant findings regarding assessment practice among the small sample. Specifically, the continued pervasive use of non-achievement assessment criteria in secondary music classrooms among responding teachers. This project found participating teachers' non-achievement assessment measures outweighed achievement-based grading practices, which was also observed by Russell & Austin (2010). As the unreliability of these non-achievement measures have

been consistently documented in music education literature, the continued use of these practices is concerning. My analysis did not find a significant relationship between school assessment culture (administrative guidance, effect of standards-based grading) and music teachers' assessment practices, which ultimately places the use of recommended grading practices in hands of mostly autonomous teachers. In that, assessments trainings for in-service teachers need more emphasis and regular attendance, as teaching experience alone has not proved to be positively correlated with the use of reliable assessment practices (Pierre & Wuttke, 2015; Russell & Austin, 2010; Zhang & Burry-Stock, 2003.)

To improve assessment practices in secondary schools, I believe we must consider the assessment curriculums of music teacher preparation programs. Aspiring music teachers should receive extensive instruction on assessment construction, which must be put into practice. Multiple opportunities to implement assessment followed by reflection of results and execution would provide valuable educational experiences for burgeoning music teachers and build upon their understanding of how to reliably assess and report grades. Assessment instruction can be presented in both music and general education contexts to promote cross-content assessment discourse, which has been found to be uncommon in schools (Cizek, Fitzgerald, & Rachor, 1995). Empowering pre-service music teachers with advanced assessment knowledge may mitigate the common use of non-achievement criteria for grading within the content, as it did for the singular case in Conway, & Jeffers (2004). Undergraduate music education programs must prioritize high levels of assessment literacy, rather than basic introductions, to ensure graduates enter the

teaching profession with confidence to reliably assess students and take pride in their grading practices.

Future studies should continue to observe the assessment practices of secondary music teachers to find trends and influences. These studies should extend their samples to include private schools, and public charters to again further examine all forms of secondary ensemble music education. Additionally, I recommend updating the Secondary School Music Assessment Questionnaire to reflect recent developments in education, particular the increased inclusion of technology. Lastly, interested scholars should more intensely focus on assessment at the classroom level by comparing the assessment formats and grading practices of secondary music teachers who exclusively grade using achievement-based criteria. These studies can determine the most preferred and possibly most effective forms of reliable assessment conducive to the music classroom to further promote achieving-based assessment practices. As the comments from this study show, music teachers see reliable assessment to be inconvenient or, at most, incompatible with secondary music education in its current form. Providing the music teachers who lack assessment confidence, are unfamiliar with reliable assessment practices, or resistant to different forms of assessment, with recommended methods and materials already use in music classroom—while deemphasizing the need for grading on attendance, attitude, and participation—will allow the field of music education to maintain similar format of instruction unique to the content while meeting the accountability expectation of the current US education climate.

Appendix A

Secondary School Music Assessment Questionnaire



Title: “Assessment in Secondary Music Classrooms: A Replication Study”

This research is being conducted by William Gonzales at the University of Maryland, College Park. We are inviting you to participate in this research project because you are a secondary school music teacher in the Mid-Atlantic region of the United States. The purpose of this research project is to investigate the assessment practices and teaching contexts of secondary school music teachers from the Mid-Atlantic region of the United States, similar purpose of the original study, Russell & Austin (2010), upon which this study is based.

The procedures involve the collection of data via a three section online questionnaire, the Secondary School Music Assessment Questionnaire (SSMAQ). Completion time for the SSMAQ is estimated to be 8 minutes. At the end of the three-week data collection window, responses will be analyzed to find possible trends and correlations.

There are no known risks from participating in this research study. There are no direct benefits from participating in this research. However We hope that, in the future, other people might benefit from this study through improved understanding of assessment practices of secondary music teachers. Any potential loss of confidentiality will be minimized by storing data files in a secure location such as a password protected flash drive. Additionally, only the primary researcher, William Gonzales, and the thesis advisor, Kenneth Elpus, will have access to these data files.

If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.

Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the investigator:

William Gonzales
wilgonzo@terpmail.umd.edu

If you have questions about your rights as a research participant or wish to report a research-related injury, please contact:

University of Maryland College Park
Institutional Review Board Office
1204 Marie Mount Hall
College Park, Maryland, 20742
E-mail: irb@umd.edu
Telephone: 301-405-0678

This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.

Clicking "I Agree" below indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study. You may print a copy of this consent form.

If you agree to participate, please click "**I Agree**" below.

I AGREE

Secondary School Music Assessment Questionnaire

Section One: School Assessment Framework

Do you provide each student with a formal grading policy?

No

Yes, verbally

Yes, in writing

What type of grades do your music ensemble students receive?

- Letter or percentage grades
- Pass/fail or satisfactory/unsatisfactory grades
- Standards-based grade reports
- No grades are assigned or required for ensembles

If you assign letter grades, approximately what percentage of your students receive each type of grade? **TOTAL SHOULD = 100%.**

A's	0 %
B's	0 %
C's	0 %
D's or F's	0 %
Total	0 %

Do music ensemble grades affect students' overall grade point averages in your school?

- No
- Yes, weighted equally with academic subjects
- Yes, but not weighted equally

Do students enrolled in an ensemble receive academic credit toward fulfillment of graduation requirements?

No

Yes

To what extent do school administrators monitor/guide the way you assess and grade students in ensembles?

- Not at all
- Somewhat
- Quite a bit
- Extensively

To what extent has the adoption of standards-based curricula by your district affected your approach to assessment and grading in ensembles?

- Not at all
- Somewhat
- Quite a bit
- Extensively
- NA – No Standards



Section Two: Specific Assessment Strategies

Indicate the approximate weight that each of the following types of criteria receives in determining an ensemble student's grade by writing the appropriate percentage in each blank. **TOTAL SHOULD = 100%**

Attendance and Punctuality	<input type="text" value="0"/>	%
Attitude (Participation, Effort, Leadership, etc.)	<input type="text" value="0"/>	%
Written Assessments of Musical Knowledge and Understanding	<input type="text" value="0"/>	%
Practice Assessments (Practice cards, student narratives)	<input type="text" value="0"/>	%
Performance Assessments of Musical Skill and Musicianship	<input type="text" value="0"/>	%
Other criteria, <i>please describe</i> : <input type="text"/>	<input type="text" value="0"/>	%
Total	<input type="text" value="0"/>	%



Answer questions below only if "Attendance and Punctuality" is part of your grading policy.

Indicate which of the following attendance criteria you consider when grading students. **SELECT ALL THAT APPLY.**

- Attendance at daily rehearsals or sectionals
- Attendance at after-school rehearsals or sectionals
- Attendance at major school concerts
- Attendance at athletic performances (i.e. pep band)
- Attendance at solo & large group festivals
- Punctuality or promptness to all rehearsals/events
- Other

Do you reduce students' grades if they have unexcused absences from major performance events?

- No
- Yes, partial letter grade reduction
- Yes, one or more letter grade reduction

Answer questions below only if "Attitude" is part of your grading policy.

Indicate which of the following attitudinal criteria you consider when grading students. **SELECT ALL THAT APPLY.**

- | | |
|--|---|
| <input type="checkbox"/> Citizenship/Courtesy/Cooperative Behavior | <input type="checkbox"/> In-class participation |
| <input type="checkbox"/> Leadership | <input type="checkbox"/> Participation in honor ensembles |
| <input type="checkbox"/> Responsibility (Bringing materials to rehearsal/concerts) | <input type="checkbox"/> Participation in support activities (clerical, fund raising) |
| <input type="checkbox"/> Instrument/Uniform care | <input type="checkbox"/> Participation in state music festivals |
| <input type="checkbox"/> Effort | <input type="checkbox"/> Other |

- Participation in private lessons

How do you assess attitudinal criteria?

- Subjective impression
- Objective documentation
- Combination of subjective & objective

Answer questions below only if "Written Assessments" are part of your grading policy.

Indicate which of the following "Written Assessments" formats you utilize when grading ensemble students. **SELECT ALL THAT APPLY.**

- | | |
|---|---|
| <input type="checkbox"/> Quizzes | <input type="checkbox"/> Major Projects/Presentations |
| <input type="checkbox"/> Major Exams | <input type="checkbox"/> Journals/Notebooks |
| <input type="checkbox"/> Worksheets | <input type="checkbox"/> Other |
| <input type="checkbox"/> Homework Assignments | <input type="text"/> |

Indicate which of the following "Written Assessment" objectives you consider when grading ensemble students. **SELECT ALL THAT APPLY.**

- Knowledge of basic terminology, symbols, or notation
 - Knowledge of music theory principles (intervals, chords, voice leading, chord progressions)
 - Knowledge of music history (style periods, composers, forms, genres, musical instruments)
 - Knowledge of compositional techniques (variation, sequence, augmentation, diminution)
 - Knowledge of cultural contexts associated with pieces of music
 - Knowledge of performance practices or pedagogical principles
 - Ability to identify musical elements by ear or sight
 - Ability to analyze and evaluate musical performances (self, section, ensemble, or other groups)
 - Ability to create small-scale original compositions or arrange existing compositions
 - Other
-

Answer questions below only if "Practice Assessments" are part of your grading policy.

Indicate which of the following types of "Practice Assessment" formats you utilize when grading ensemble students. **SELECT ALL THAT APPLY.**

- Quantitative Reports/Practice Cards (documenting amount of time students practice)
 - Qualitative Reports (describing what and how students practice)
 - Practice tapes (recordings of student practice sessions)
 - Other
-

Answer questions below only if "Performance Assessments" are part of your grading policy.

Indicate which of the following types of "Performance Assessment" formats you utilize when grading ensemble students. **SELECT ALL THAT APPLY.**

- | | |
|---|--|
| <input type="checkbox"/> Playing exams, live & in-class | <input type="checkbox"/> Sectional performances in class |
| <input type="checkbox"/> Playing exams, live but outside of class | <input type="checkbox"/> Ensemble performances in concert |
| <input type="checkbox"/> Playing exams, audiotaped | <input type="checkbox"/> Ratings at solo & ensemble festival |
| <input type="checkbox"/> Playing exams, videotaped | <input type="checkbox"/> Ratings at large-group festival |
| <input type="checkbox"/> Auditions | <input type="checkbox"/> Other |
| | <input type="text"/> |
| <input type="checkbox"/> Chair challenges | |

Indicate which of the following "Performance Assessment" objectives you consider when grading band students. **SELECT ALL THAT APPLY.**

- Performance technique (scales, etudes, etc.)
 - Student-prepared performance of band music excerpts (tone, accuracy, & musicality)
 - Student-prepared performance of other material or repertoire (tone, accuracy, & musicality)
 - Sight-reading performance
 - Improvised performance
 - Memorized performance
 - Other
-

Which of the following approaches do you use in conjunction with performance assessments? **SELECT ALL THAT APPLY.**

- Global impression
 - Checklists
 - Rating Scales
 - Rubrics
 - Other
-

What other comments or concerns do you have about assessing and grading ensemble students?

Section Three: Teacher Background and Teaching Context

Which do you consider your primary teaching assignment?

Band

Choir

Orchestra

Which do you consider your primary teaching level?

Middle School/Junior High

High School

In what state do you teach?

On average, how many minutes of instructional time per week do you have with each of your ensemble classes?

minutes per week

What is the total number of ensemble students you are responsible for assessing and grading?

students

How many ensemble performances are you responsible for in a typical school year?

performances

How confident are you in your ability to properly assess and grade ensemble students?

- Not confident at all
- Somewhat confident
- Very confident
- Extremely confident

Including the current academic year, how many years of experience do you have?

years teaching music (all areas and levels)

What is your highest earned degree?

Bachelors

Masters

Doctorate

In which of the following settings have you received additional training in assessment techniques? **SELECT ALL THAT APPLY.**

- Conference Clinics
- District In-services
- University Workshops
- Graduate Courses
- Other

Are you willing to answer a few additional questions about your experiences with assessment & grading?

No

Yes

References

- Abeles, H. F. (2010). Assessing music learning. *Critical issues in music education: contemporary theory and practice*, 167-193.
- Abril, C. R., & Gault, B. M. (2008). The state of music in secondary schools the principal's perspective. *Journal of Research in Music Education*, 56(1), 68-81.
- Asmus, E. P. (1999). Music assessment concepts. *Music Educator's Journal*, 86(2), 19-24.
- Bradley, D. (2015). Hidden in plain sight: Race and racism in music education. In C. Benedict, P. Schmidt, G. Spruce, & P. Woodford (Eds.), *The Oxford handbook of social justice in music education*, (pp. 190-203). New York, NY: Oxford University Press.
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159-179.
- Conway, C., & Jeffers, T. (2004). Parent, student, and teacher perceptions of assessment procedures in beginning instrumental music. *Bulletin of the Council for Research in Music Education*, 16-25.
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative research*. Upper Saddle River, NJ: Prentice Hall.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53-72.

- DeLuca, C., & Bellara, A. (2013). The current state of assessment education aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, doi: 0022487113488144.
- Elpus, K. (2013). Is it the music or is it selection bias? A nationwide analysis of music and non-music students' SAT scores. *Journal of Research in Music Education*, doi: 0022429413485601.
- Elpus, K. (2014). Evaluating the effect of No Child Left Behind on US music course enrollments. *Journal of Research in Music Education*, doi: 0022429414530759.
- Elpus, K., & Abril, C. R. (2011). High school music ensemble students in the united states a demographic profile. *Journal of Research in Music Education*, 59(2), 128-145.
- Fautley, M. (2015) Music education assessment and social justice: Resisting hegemony through formative assessment. C. Benedict, P. Schmidt, G. Spruce, & P. Woodford (Eds.), *The Oxford handbook of social justice in music education*, (pp. 513-524). New York, NY: Oxford University Press.
- Fautley, M., & Colwell, R. (2012). Assessment in the secondary music classroom.
- Hanna, W. (2007). The new Bloom's taxonomy: Implications for music education. *Arts Education Policy Review*, 108(4), 7-16.
- Hoffman, A. R. (2013). Compelling questions about music, education, and socioeconomic status. *Music Educators Journal*, 100(1), 63-68.
- Johnson, R. (2008). Overcoming resistance to achievement-based unit grading in secondary physical education. *Journal of Physical Education, Recreation & Dance*, 79(4), 46-49.

- Kimpton, P., & Harnisch, D. L. (2008). *Scale your way to music assessment: The ultimate guide to creating a quality music program*. Chicago, IL: GIA Publications.
- Kimpton, P., & Kimpton, A. K. (2013). *Grading for musical excellence: Making music an essential part of your grades*. Chicago, IL: GIA Publications.
- Kimpton, P., & Kimpton, A. K. (2014). *Common Core: Re-imagining the music rehearsal and classroom: Standards, curriculum, assessment, instruction*. Chicago, IL: GIA Publications.
- Kotora Jr, E. J. (2005). Assessment practices in the choral music classroom: A survey of Ohio high school choral music teachers and college choral methods professors. *Contributions to Music Education*, 65-80.
- Mackey, A. (2012). Why (or why not), when and how to replicate research. *Replication research in applied linguistics*, 2146.
- McCoy, C. W. (1988). An exploratory study of grading criteria among selected Ohio ensemble directors. *Contributions to Music Education*, (15), 15-19.
- McCoy, C. W. (1991). Grading students in performing groups: A comparison of principals' recommendations with directors' practices. *Journal of Research in Music Education*, 39(3), 181-190.
- Mello, J. A. (2010). The good, the bad and the controversial: the practicalities and pitfalls of the grading of class participation. *Academy of Educational Leadership Journal*, 14(1), 77.

- Odegaard, D. (2009). *Music curriculum writing 101: Assistance with standards-based music curriculum and assessment writing: For band, choir, orchestra, and general music*. Chicago, IL: GIA Publications.
- Pellegrino, K., Conway, C. M., & Russell, J. A. (2015). Assessment in Performance-Based Secondary Music Classes. *Music Educators Journal*, 102(1), 48-55.
- Pierre, N. A. S., & Wuttke, B. C. (2015). Standards-Based Grading Practices Among Practicing Music Educators Prevalence and Rationale. *Update: Applications of Research in Music Education*, doi: 8755123315604468.
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372-1380.
- Reynolds, C. R., Livingston, R. B., Willson, V. L., & Willson, V. (2010). *Measurement and assessment in education*. Upper Saddle River, NJ: Pearson Education International.
- Russell, J. A., & Austin, J. R. (2010). Assessment practices of secondary music teachers. *Journal of Research in Music Education*, 58(1), 37-54.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90.
- Weiner, B. (1972). Attribution theory, achievement motivation, and the educational process. *Review of Educational Research*, 42(2), 203-215.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342.