ASSESSMENT OF OHIO MUSIC TEACHERS:
CHALLENGES AND IMPLICATIONS

Thesis
Submitted to
The School of Education and Allied Professions of the
UNIVERSITY OF DAYTON

In Partial Fulfillment of the Requirements for
The Degree of
Master of Science in Education

By
Melissa Anne Durst
Dayton, Ohio
May 2012
ASSESSMENT OF OHIO MUSIC TEACHERS: CHALLENGES AND IMPLICATIONS

Name: Durst, Melissa Anne

APPROVED BY:

______________________________________________________
Linda A. Hartley, Ph.D.
Committee Chair

______________________________________________________
Diana M. Hunn, Ph.D.
Committee Member

______________________________________________________
Nancy E. Parks, Ed.D.
Committee Member
ABSTRACT

ASSESSMENT OF OHIO MUSIC TEACHERS: CHALLENGES AND IMPLICATIONS

Name: Durst, Melissa Anne
University of Dayton

Advisor: Dr. Linda A. Hartley

Legislative mandates such as No Child Left Behind and Race to the Top have shaped an educational paradigm shift- effective teaching is defined by demonstrating student growth within a content area. Data is gathered and analyzed yearly, with results potentially impacting teacher evaluations, teacher pay, and hiring and firing decisions.

Standardized testing is one of the most common ways of assessing students. Test results give clear and coherent views of student growth over a specific period of time, but rely heavily on quantitative measures. There are numerous other means to assess student growth however, and to achieve the most efficient and effective evaluation, multiple assessments should be considered that consist of qualitative measures as well.

Non-tested subjects, such as music, are also under the same requirements to measure student growth to evaluate teacher effectiveness. There is no national or state-wide assessment that is currently recognized and accepted for measuring music student
growth. Music teachers are responsible for teaching “knowledge” - a belief or consensus of fact; and “skills” - the aptitude for performing a specific task.

This thesis specifically examines the following questions:

1. How should the growth of music students in the state of Ohio be effectively assessed?; and
2. How should the effectiveness of music teachers in the state of Ohio be determined?

Utilizing multiple, standards-based assessments, such as portfolios, performances, observations, and pretest/posttest models, student growth in music can be effectively and efficiently measured for music students in the state of Ohio. Music teacher assessments in the state of Ohio should include evidence of a music teacher’s capacity to plan lessons, establish a healthy learning environment, utilize best teaching practices, and create and maintain opportunities for professional growth. A new evaluation tool was created for assessing music teachers. This tool, created after reviewing available literature, establishes four domains of teaching, similar to the Charlotte Danielson domains, and conforms to recent Ohio legislative mandates.
ACKNOWLEDGEMENTS

I am indebted to many people who encouraged me to follow my dreams and complete this thesis, which I hope is the first of many academic publications. I would like to acknowledge my friends and coworkers at Centerville City Schools for their insights to teaching in and out of the classroom. Many thanks also to my non-teaching friends and family for the extra set of eyes for proofreading, ensuring that this work is accessible to a wider audience. A special thanks to Blake Stamm who was instrumental in helping with the day to day tasks and ensured that I would be taken care of, even when I was in “thesis mode.”

And last but not least, I am most indebted to Dr. Linda Hartley. It is through her guidance and evaluation that this is even possible. Dr. Hartley has been influential in my career throughout my undergraduate degree and through my masters. Undoubtedly, I would not be here without her.
TABLE OF CONTENTS

ABSTRACT ....................................................................................................................................... iv
ACKNOWLEDGEMENTS................................................................................................................ vi
LIST OF FIGURES....................................................................................................................... viii
DEFINITIONS OF TERMINOLOGY............................................................................................ ix
CHAPTER 1: INTRODUCTION...................................................................................................... 1
CHAPTER 2: REVIEW OF RELATED LITERATURE...................................................................... 3
CHAPTER 3: METHODOLOGY AND RESULTS .......................................................................... 20
CHAPTER 4: DISCUSSION........................................................................................................... 30
CHAPTER 5: EDUCATIONAL IMPLICATIONS.......................................................................... 35
REFERENCES................................................................................................................................. 38
LIST OF FIGURES

FIGURE 1: Proposed student growth measures .........................................................24
FIGURE 2: Domain Overlaps......................................................................................28
FIGURE 3: Assessment for Music Teachers in the State of Ohio (Proposed).................29
DEFINITIONS OF TERMINOLOGY

**Assessment** – “any systematic method of obtaining information about knowledge and skills and using it to draw inferences about people or programs.” (Tung, 2010, p. 50)

Assessment scores are gathered over a span of time before results are analyzed.

**Evaluation** – method of measuring and/or observing knowledge or skills. Evaluations should be viewed as a snapshot in time, monitoring the knowledge that is known at a specific point in time. Evaluation results are often compared to a baseline statistic, such as a percentage.

**Local Education Authority (LEA)** – committee of people that set forth policy for a district or school.

**National Assessment of Educational Progress (NAEP)** – “congressionally authorized project of the National Center for Education Statistics...within the Department of Education” (Keiper, Sandene, Persky, & Kuang, 2009, p. 39). NAEP assessments have been conducted in various content areas including reading, math, science, and the arts. Schools are randomly assigned to participate in the test. The music version of the NAEP is a pencil-and-paper test gauging student knowledge on a variety of musical topics, including music theory, music history, and aural identification. Students answer multiple choice or short response questions.
National Board for Professional Teaching Standards (NBPTS) – a national accreditation process for teachers. This extensive process relies on a combination of classroom artifacts, portfolio, classroom observation, teacher knowledge, and teacher reflection.

No Child Left Behind (NCLB) – educational reform policy signed into law in 2001 by President George W. Bush. This policy is one of the first that specifically places the responsibility of student growth on teachers, administrators, and individual schools. While controversial, NCLB also highlighted where schools might not have fully served the entire student population.

Ohio Teacher Evaluation System (OTES) – system of evaluating teachers in the state of Ohio. As of 2012, the current proposal for OTES is based 50% on student growth (measurable objectives) and 50% teacher performance (including professionalism, learning environment, etc). Teachers will be given a rating and a teacher’s salary will depend upon their rating.

Performance Assessment- “product- and behavior-based measurements based on setting designed to emulate real-life contexts or conditions in which specific knowledge or skills are actually applied” (American Educational Research Association, American Psychological Association National Council on Measurement in Education, 1999; as cited in Tung, 2010).

Race to the Top – “a competitive grant program to encourage and reward States that are implementing significant reforms in the four education areas described in the ARRA (American Recovery and Reinvestment Act of 2009); enhancing standards and assessments, improving the collection and use of data, increasing teacher effectiveness.
and achieving equity in teacher distribution, and turning around struggling schools” (U.S. Department of Education, 2010, p. 3).

**Student Growth**- “the change in student achievement between two or more points in time” (U.S. Department of Education, 2010, p. 21).

**Tennessee Value-Added Assessment System (TVAAS)** – educational system that analyzes student test score data and estimates the effects of individual teachers on score gains” (Kupermintz, 2003, p. 287); developed in the 1980s by Dr. William L. Sanders at the University of Tennessee.

**Value-Added** – a statistical representation of academic growth within a set time frame, based upon student performance on a standardized test.
CHAPTER 1
INTRODUCTION

At the turn of the millennium, a paradigm shift rocked the education profession. Terms like “overhaul” and “reform” were often paired with “education.” Federal legislative actions such as No Child Left Behind Act of 2001, signed into law by President George W. Bush, and Race to the Top, by President Barack Obama in 2009, require greater accountability from individual teachers to increase student achievement. Teacher effectiveness has taken high priority in education reform conversations. Since student growth is the desired result, teachers are held more accountable for student learning and growth.

There are two variables to the learning equation of teacher effectiveness and differences in student learning. In 2003 Kupermintz offered the following concerning the two variables, “[t]o claim that teacher effectiveness is the cause of student score gains is at best a necessary, trivial truth similar to the observation that „all bachelors are unmarried”’” (p. 289). The relationship between student and teacher is precarious as well as symbiotic. Students need teachers to gain the skills needed for future success, and legislators and administrators are placing more importance and weight upon the success of students within the classroom. “One key goal of education is to give students the skills needed to ensure a productive career and sustain their economic well-being” (Lavy, 2007, p.88).
Skills and knowledge can be defined from two different perspectives. For the purposes of this paper, “skill” is defined as the aptitude for performing a specific task; “knowledge” is a belief or consensus of fact. State standardized testing is one popular means to assess student growth, however “growth” has a narrow definition as the change of student knowledge from one year to the next and little challenge to student skill.

While it may appear uncomplicated to assess student growth and teacher effectiveness in some content areas, other content areas struggle to create effective tools to measure student growth within the classroom. Non-tested subjects, including visual arts, physical education, foreign language, social studies/civics, and music, are left with the same legislative mandates, but without previously established evaluation techniques. Creating appropriate assessments for students in non-tested subjects is a current area of discussion and initiative. Various organizations, including the National Association for Music Education, released a teacher evaluation position statements, outlining suggested criteria for measuring student achievement with regards to teacher evaluation (NAfME, n.d.). Other states, including New York, are creating their own data driven music assessment tools (New York State Education Department, 2011).

Music teachers are in a precarious situation; they must devise the means to quantify and qualify everything that their students are doing within the classroom and rehearsal space, since there is no state or national standardized assessment for music students. This thesis specifically examines the following questions:

1) How should the growth of music students in the state of Ohio be effectively assessed? and

2) How should the effectiveness of teachers in the state of Ohio be determined?
CHAPTER 2
REVIEW OF RELATED LITERATURE

Assessment & Student Data

Standardized testing is one of the most common ways of assessing students. Its means are relatively simple and predetermined based upon state standards. Students are given an opportunity to express correct answers through multiple-choice selections or short response answers. Tests like the Ohio Achievement Assessment occur at the same time frame throughout the state and the same questions are asked of all students. The tests cover materials established through state standards. There is little room for bias or skewing of scores (aside from unethical behaviors). Since the tests are standard throughout a specific geographical area (e.g. state, district, county) it is relatively easy to look at student scores and subsets to compare the growth of students and the effectiveness of their teachers.

One of the main concerns with utilizing only test scores to assess the growth of students is that these tests merely offer a snapshot into a student’s potential. “Test scores are limited in the information they can provide” (Little, Goe, & Bell, 2009, p. 2). There are numerous variables that can affect a student’s performance on a test. Consider Maslov’s hierarchy of needs. If students’ needs (physical, emotional, or emotional) are not met, then students will not be able to perform to their full potential. Students enter
classrooms sometimes with a great deal of emotional baggage, and there is no possibility to quantify or assess the toll that life takes upon an individual.

For music educators, one of the most challenging aspects of the assessment conversation is the lack of an established and accepted standardized test.

Race to the Top criteria specify using both the National Assessment of Educational Progress (NAEP) and the assessments required under section 1111 (b)(3) of the Elementary and Secondary Act of 1965, as amended (ESEA) to track a State’s increases in student achievement and decreases in the achievement gap over the course of the State’s grant. (U.S. Department of Education, 2010, p. 16)

In 2008, NAEP administered a standardized test for the arts to nearly 7,900 students in public and private schools; half of the students took the test for visual arts, the other half took the test for music. The NAEP test in music established criteria for listening to music, identifying historical contexts of works, identifying instruments, and writing music within a theoretical context, yet there was no assessment over performing music.

Alternative to a standardized test like the NAEP, a performance assessment could also accurately evaluate students by requiring them to “create an original answer or product; use higher order thinking and 21st Century skills; demonstrate thinking processes; evaluate real world situations” (Tung, 2010, p. 2). Performance assessments ask for students to go beyond the multiple-choice answers. Performance assessments are not relegated to formative or summative learning, nor are they bound by time frames.

Performance assessments provide students with more ways to show what they know and can do, allow students with different learning styles more opportunities to succeed, and engage students more in their own learning and interests, because
they include reflection and demonstration of the thinking process. Performance assessments are more closely aligned with real world skills that students will need. (Tung, 2010, p. 4)

Within a music context, performance assessments could include everything from the historical, theoretical, and performance bases. Students might constantly be performing concerts or musicals or recitals, which suits performance assessments well. Performance assessments are defined as “product- and behavior-based measurements based on settings designed to emulate real-life contexts of conditions in which specific knowledge or skills are actually applied” (American Educational Research Association as cited in Tung, 2010, p. 2)

**Student Growth**

The issue at heart with legislative mandates is to demonstrate just how teachers are effective, how precisely students benefit from this individual as a teacher in their lives and how our students are learning. “Student growth is defined generally as the change in student achievement between two or more points in time” (U.S. Department of Education, 2010, p. 22) In Colorado, student growth is compared to their academic peers; if students are out-performing the average of their peers, students are performing well (Goe & Holdheide, 2011).

Value-added models are popular measurements of illustrating student growth. In a value-added model, it is believed that “students” prior achievement on standardized tests can be used to predict their achievement in a specific subject the next year” (Little, Goe, & Bell, 2009, p. 4). Value-added is intended to be an objective view of teacher effectiveness on student growth.
Measuring student growth is easily done through pre-tests and post-tests, comparing the score difference, and establishing growth. Growth in a music class can come in a variety of forms, from acquiring basic knowledge to improvement in performance skills. Performance skills are cumulative; students must first learn how to produce a quality tone on their instrument before they could attempt a difficult solo or symphonic work. Comparing a first rehearsal to a performance, one can expect that there has been marked improvement. However the question remains- did the teacher effectively teach her students, or was class a series of repetitions to improve technique?

**Teacher Effectiveness**

It truly takes a village to raise a child. The responsibility of educating students lies not only within teachers, but parents, districts, public officials, and students (Little, Goe, & Bell, 2009; Heubery & Hauser as cited in Little, Goe, & Bell, 2009). Even though there are numerous factors outside a teacher’s control, there are many things that teachers can do to ensure that their students are achieving at a high level.

Race to the Top defined an *effective* teacher as one whose students achieved at least one grade level of academic growth during the course of the year and a *highly effective* teacher as a teacher whose students achieved at least one and a half grade levels of academic growth during that time frame. (Goe & Holdeheide, 2011, p. 5)

Little, Bell, and Goe (2009) selected five points that make an effective teacher: holding all students to high expectations; creating a positive environment for students; using a variety of resources for educational opportunities; establishing a fair and safe
environment for students; and collaborating with other professionals to ensure student success.

With a potential for collaboration and the influence of family and society, it is difficult to distinguish the benefits that one student reaps from one teacher versus another teacher. Collaboration and team-teaching is not only a reality of the education profession but an expectation. “Identifying precisely what one teacher contributes to a student’s performance and separating his contribution from those of other teachers, the school, the principal, and the family is extremely difficult” (Lavy 2007, pp 91-92). Students enter their classes with a variety of prior knowledge at various stages of development (physically, academically, emotionally, etc).

Isolating the teaching of one person can be difficult. Some students might have the opportunity to seek enrichment, thus increasing their knowledge or aptitude. When students have the opportunity to participate in music ensembles, lessons, or other activities outside the music classroom, it is nearly impossible to effectively gauge where one experience leaves off and the other picks up. It is most ideal for the student that the two experiences (in and outside of the classroom) enhance the student’s understanding of music. However, assessing the growth of this particular student is difficult. It must be considered that not all students are able to take advantage of the same opportunities outside of the school day, or their activities should be scaffolded differently in the classroom. A differentiated classroom approach is necessary to meet each student at his knowledge and skill level; music teachers must recognize the diversity of knowledge in their classrooms.
A student’s aptitude or work ethic can also impact teacher effectiveness, if based on value-added means. “Equally competent teachers will produce different results with groups of student that differ appreciably in cognitive, affective, and motivational aptitude profiles” (Kupermintz, 2003, p. 291). Students respond differently to different teachers; often a student who “loves” his or her teacher, possibly exerts a greater effort to please the teacher. Sometimes students who are more apathetic to their teachers pass along the apathy to their work. The student could possess the skills and aptitude needed for evaluations, but often does not try.

What can a teacher can do if students do not or cannot do their best work?

Students should be held to a greater accountability if teachers are to be held to a greater accountability. Publishing test scores appear to hold little sway to motivate or deter students. Yet, “teacher effectiveness is often defined as the ability to produce gains in student achievement scores” (Little, Goe, & Bell, 2009, p. 1). Teachers are tasked with establishing concrete evidence supporting that their teaching had a positive impact on students to improve their scores, regardless of individual student concerns.

**National Legislation**

The most recent legislative education action handed down on a federal level came in the form of the American Recovery and Reinvestment Act of 2009 (ARRA). Within President Obama’s plans for encouraging economic growth came an opportunity for States to overhaul their education standards and programs at a chance for substantial federal stimulus monies. Race to the Top is a:

competitive grant program to encourage and reward States that are implementing significant reforms in the four education areas described in the ARRA: enhancing
standards and assessments, improving the collection and use of data, increasing teacher effectiveness and achieving equity in teacher distribution, and turning around struggling schools. (U.S. Department of Education, 2010, p. 3)

Race to the Top winners must follow through with the promises in their proposals, however States can apply for amendments. Race to the Top established three requirements for student learning: rigorous, between two points in time, and comparable across classrooms (Goe & Holdheide, 2011).

This legislative action came on the heels of No Child Left Behind (NCLB), signed into law by President George W. Bush. NCLB placed a greater pressure for schools and districts to ensure that all children were receiving a quality education by publishing individual school report cards, documenting the performance of various subgroups based on race or socioeconomic status. NCLB also pushed for 100% passage rate in reading. Individual states were responsible for establishing tested content and at what level was considered “passing.” This pressure could lead states to alter testing content, without altering instruction, or other means to “game” this system.

Ohio Legislation

Ohio applied for the Race to the Top competitive grant, and was awarded within the Phase 2 range not to exceed $400 million (U.S. Department of Education, 2010). In accordance with Race to the Top funding, educational requirements were written into the Executive Budget for Fiscal Years 2012 and 2013 in the state of Ohio, “the State Board of Education has identified the following three objectives: teach 21st Century knowledge and skills for real-world success, effectively deliver support for a high quality education,
and provide sufficient resources that are effectively managed” (Ohio Office of Budget
and Management, 2011).

Ohio teachers are asked to ensure that their students would have 21st Century
skills such as critical thinking; communication; collaboration; creativity; innovation;
information, media, and technology literacy; flexibility; initiative; social and cross
cultural skills; productivity and accountability; and leadership (Partnership for 21st
Century Skills, 2010).

These skills naturally coincide with opportunities in the music. Perhaps a unit on
improvisation in which the students are asked to work in small groups to improvise a
short rhythm pattern or melody (communication, collaboration, creativity, innovation,
leadership). Students could also listen and identify improvised sections within popular
music (critical thinking, media literacy). In a performance ensemble, students are often
given specific tasks as a section leader to set a good example for the other members in a
section. Learning how to work together to make music is a crucial skill for ensemble
performance. Listening to works performed by professional and non-professional groups
from around the world offers an opportunity to dialogue not just about music
performance, but the impact of music on various cultures.

In 2011, the State Board of Education established the Ohio Teacher Evaluation
System (OTES). Under this system, half of a teacher’s evaluation is based on student
growth, the other half is based in teacher performance. Student growth is assessed at
above growth, expected growth, or below growth. Teacher performance is assessed
through a four point scale, four as the highest and one as the lowest. Combining the two
scores, teachers receive an overall designation of Accomplished, Proficient, Developing,
or Ineffective (Ohio Department of Education, 2011). If student growth is above expectation and a teacher’s performance is evaluated as a three or four, the teacher’s overall evaluation is listed as “Accomplished.” If student growth meets expectations and the teacher’s performance is assessed as a two, the overall evaluation is listed as “Developing.” If student growth is below expectation and a teacher’s performance is listed as a one or two, the teacher’s overall evaluation is listed as “Ineffective.” If a teacher is evaluated as “Ineffective,” it is possible that the teacher would have to implement an improvement plan and reevaluation or face dismissal.

**Teacher Evaluation**

For an effective teacher evaluation, research supports gathering data from multiple sources over a period of time. (Little, Goe, & Bell, 2009; Kupermintz, 2003). A teacher evaluation is something that must be carefully weighed, either leading to progressing further in the field, increased pay, or dismissal. Data can be collected and analyzed in a variety of ways, and depending on analysis, numerous conclusions can be drawn. For an effective evaluation, there should be a clear understanding of what the evaluation should encompass, including how the data gathered will be analyzed (Little, Goe, & Bell, 2009).

There are some serious concerns with some teacher evaluation processes. There must be some weight and balance to the system or else

an outstanding teacher who taught more students will be correctly identified, whereas an equally remarkable teacher serving a more transient student population would appear less exemplary because her performance score will be
pulled towards the system average... a poor teacher may evade detection.

(Kupermintz, 2003, p. 291)

There needs to be a broad enough evaluation structure that allows for teacher evaluations to consider factors beyond their control. Adequate yearly progress (AYP) was implemented to ensure that students were progressing along a statically appropriate track. However, statistics cannot encompass the diversity within a classroom, school, or family environments. AYP is student specific and measures the growth of student knowledge within a content area, not applicable skills (including 21st Century skills that are now so highly sought).

Adequate yearly progress is a form of absolute criterion, comparing individual student scores to a predetermined threshold, and there is no official scale of AYP specifically with music. Student growth could possibly be measured by relative criteria, comparing class scores to other teachers within the geographic area, content, and grade levels. This would allow for a more equitable sharing of information (comparing apples to apples). “Teachers should be evaluated on the basis of their true performance, not random variation in performance” (Lavy, 2007, p. 100).

There are many factors that go into the concept of an “effective teacher.” Knowledge of content and the ability to pass along knowledge to students seem to encompass the clinical definition. According to the U.S. Department of Education:

The definition of effective teacher, [has] been clarified... schools must use multiple measures in determining effectiveness... Effectiveness must be evaluated, in significant part, based on student growth, and... provide examples of
supplemental measures a State or LEA (local education authority) may use.

(2010, p. 21)

Within this definition, states are able to create their own means of defining “effectiveness, however, part of that process must include student growth. This creates an opportunity for states to create an evaluation that, in part, incorporates application of 21st Century skills in the classroom, and the five points defined by Little, Goe, and Bell (2009).

Charlotte Danielson developed a “framework for teaching” that encompassed four domains of teaching: planning and preparation; classroom environment; instruction; and professional responsibilities (2007). Evaluators can relatively easily follow these four domains that can lead to effective teaching. States, including Ohio, model their teacher evaluation after Danielson’s framework (Ohio 2011).

Examples of Student and Teacher Evaluation Models

In the 1980, Dr. William Sanders created an “educational system that analyzes student test score data and estimates the effects of individual teachers on score gains” (Kupermintz, 2003, p. 287). This system is commonly referred as the Tennessee Value-Added Assessment System (TVAAS). Recognizing that it would be difficult to determine an individual teacher’s effect on a student’s growth, the TVAAS model utilizes three different statistical models:

(a) a system model estimating average performance [of] a particular school [district], for each year, grade, and academic subject, (b) a school model estimating average performance for a particular school within a [district], and (c)
a teacher model estimating the average student performance associated with a particular teacher in the [district]. (Kupermintz, 2003, pp. 289-290)

The system and school models could lead to the risk of inequality in the workplace, knowing that some teachers will go above and beyond to pick up whatever “slack” is felt (Goe & Holdheide, 2011; Kupermintz, 2003; Lavy, 2007). These models also lend themselves to potential peer pressure from coworkers to alter teaching techniques or content coverage, and a power hierarchy could ensue. The teacher model could lead to a disinterest in collaboration, since every teacher is for himself/herself.

**Value-added assessments.** Value-added measures evaluate student produced data, looking for growth. “Student growth, not raw student achievement or proficiency data is therefore the relevant measure on which to focus teacher and principal evaluations” (U.S. Department of Education, 2010, p. 21). Lavy argues that, “value-added measures, in particular, are more appropriate for measuring individual teacher effectiveness and ensuring fair ranking” (p. 101).

The National Assessment of Educational Progress (NAEP) administered a test specifically for the arts in America in 2008. Students were asked to respond to music in a variety of means, from short, free response and multiple-choice questions. Questions drew upon a diverse amount of knowledge, such as aurally identifying instruments, score reading, identifying the origin of music, and music history. Only one question assessed a student’s skill to create music. According to *The Nation’s Report Card: Arts 2008 Music & Visual Arts*, budget constraints prevented further assessments based on skill (2009, p. 9). This assessment isn’t a true value-added assessment; it is administered to random schools, public and private, and it is currently not possible to link student data to the test.
However, the NAEP test is the closest assessment music teachers have to a value-added, pencil-and-paper test.

Value-added measures can be costly to establish a valid test, score, and analyze, especially if this is the first time that value-added measures are used. Value-added measures should encompass more than just comparing test scores. Not all value-added tests can effectively assess a student’s growth. It is important to evaluate all aspects of a student’s growth and development. Otherwise, that could create an unfair culture within a classroom between the students who are high performing versus the low performing (Kupermintz, 2003; Lavy, 2007). All students deserve to be challenged at their appropriate developmental level. Lavy (2007) suggests an adjustment for factors outside the teacher’s control but not at the sacrifice of the transparency that a basic value-added model can provide. Value-added models, paired with other models, could be an effective means to evaluate teachers.

**Classroom observations/principal observations.** Classroom observations are a traditional model of teacher evaluation. A trained evaluator, often a principal, observes a class, and meets with the teacher to discuss the observations and offer suggestions for improvement. “Observations are considered the most direct way to measure teaching practice because the evaluator can see the full dynamic of the classroom” (Little, Goe, & Bell, 2009, p. 7), especially when the observer is closely connected to the school or content area. However, little information is offered concerning actual student achievement (Goe & Holdheide, 2011).

Principals have a unique perspective on classroom observation. The staff, student, and community are areas of familiarity, and the principal is often a well
respected community figure. However, they juggle the duties of congratulating student or staff achievement and issuing student or staff discipline notices, creating professional development opportunities, and serving as a liaison between school and district, state, and federal organizations. Burdening principals with additional requirements for teacher assessments will not necessarily create an efficient and effective assessment. Training principals to be effective evaluators can take away time from other valuable tasks. One study (Brandt et al. as cited in Little, Goe, & Bell, 2009, p. 8) found that only 8 percent of districts mentioned evaluation training as a component of their teacher evaluation systems... although most evaluations were being used for high-stakes, summative purposes, there was little evidence that they were being used in a reliable and valid manner.

Bias could be a concern with observations as some evaluators intentionally find the evaluation materials they are looking for and disregard other evidence.

Within a music classroom, there are specific techniques and procedures that a music specialist would be able to correctly identify as part of pedagogy. It is most ideal to have an observer with some training, not just in evaluating teachers, but also in the content taught. Thus, evaluators can offer constructive critiques to improve music instruction.

**Analysis of classroom artifacts.** The individual classroom is a wealth of knowledge about teachers and how efficiently they carry out lessons. Classroom artifacts such as “lesson plans, teacher assignments, assessments, scoring rubrics, student work, and other artifacts... [are used to] glean a better understanding of how a teacher creates learning opportunities for students on a day-to-day basis” (Little, Goe, & Bell, 2009, p.
9). Analyses of classroom artifacts are found to be a fairly reliable way to evaluate a teacher’s effectiveness. This assessment is “practical and feasible because the artifacts have already been created by teachers, and the procedures do not appear to place unreasonable burdens on teachers” (p. 10). To ensure reliability, great effort must be placed in preparing scorers; it could benefit teacher and scorer if the scorer was also knowledgeable in the content being evaluated. Paired with a classroom observation, evaluators would have a whole perspective of the artifacts in use.

In addition to lesson plans, visual aids, and other artifacts, a performance portfolio serves as a comprehensive view of student and/or teacher work. “Portfolios are a collection of materials compiled by teachers to exhibit evidence of their teaching practices, school activities, and student progress” (Little, Goe, & Bell, 2009, p. 10). Portfolios might seem similar to gathering classroom artifacts, but a portfolio is created by a teacher with the purpose of the specific evaluation, whereas the classroom artifacts are created as part of student evaluation. Depending on the requirements of the portfolio, classroom artifacts might be required.

Music teachers can create classroom performance portfolios. Creating music is a unique opportunity for students. Performances can happen within the classroom, through a school concert or recital, or musical theatre performance. These performances can be recorded or reflected upon for students to articulate their self-assessment of growth. Reviewing the performance portfolio could demonstrate student growth.

Portfolios are a popular method to evaluate a teacher’s effectiveness as they can demonstrate efforts inside and outside the classroom, readily adaptable to any content area and any grade level. Also, assembling a portfolio can be an enormous time burden
for teachers. National Board for Professional Teaching Standards (NBPTS) certification is partially based on a portfolio review. A National Board Certification is often seen as a feather in the cap for educators, a pay increase in some schools, and in Ohio, can lead to Lead Professional Educators License. According to Little, Goe, and Bell, “teaching portfolio studies have not produced conclusive results about their reliability or validity in measuring teacher effectiveness.” (2009, p. 10)

Included in classroom artifacts could be a self-report of practice, where teachers document the standards and efforts achieved within their own classrooms. Self-reports of practice are diverse, and their value depends on the purpose and use of such reports. Self reports can include surveys, checklists, teaching logs, rating systems, or frequency of practice standards. “Teachers are the only ones with full knowledge of their abilities, classroom context, and curricular content, and they can provide insights than an outside observer might not recognize” (Little, Goe, & Bell, 2009, p. 12).

**Student survey.** Popular websites and students rate secondary teachers on a scale, often a Likert scale. However there are many concerns with student surveys as a sole indicator of teacher effectiveness. Students sometimes correlate student achievement with effective teaching, but do not take into consideration other indicators, such as classroom management, curriculum and content knowledge, and professional development. Student surveys might be appealing as part of a teacher’s evaluation as these evaluations are inexpensive and time efficient (Little, Goe & Bell, 2009).

**Summary of Findings**

Due to the changing political climate on a federal and state level, teachers are being held more accountable to their responsibilities of imparting content knowledge to
their students. This accountability materializes as proof of student growth. The most effective teachers, hypothetically, will demonstrate the largest gains in student growth. While many content areas, such as reading and mathematics, have an established assessment in the form of a standardized test, non-tested areas, such as music, are left to create their own assessments, establishing student growth for their own classrooms.

There are numerous ways to establish student growth, without utilizing a standardized test. Items such as student artifacts, classroom observations, principal evaluations, portfolios, classroom artifacts, self-reports of teaching practices, and surveys can effectively be used collaboratively for the common goal of establishing music teacher effectiveness. As Lavy stated in 2007, “No single system is likely to capture all important behaviors. One way to avoid this ... is to include two outcomes, for instance, test scores, and processes... and teacher practices” (p. 101). Schools, districts, and state and federal policymakers must understand that there should not be only one measure for student growth and teacher effectiveness. As Goe and Holdheide found:

Waiting until the measures are perfected may be impractical... the measures may be weak evidence of validity in the first attempts at implementation, states and districts will benefit from creating a process to continually evaluate and strengthen the measure or eliminate those that continue to show weak evidence of validity. (2011, p. 10)
CHAPTER 3
METHODOLOGY AND RESULTS

To create a tool useful to music teachers and administrators, a mixed-method research process was employed. This study is neither truly quantitative, “research in which numeric data are collected and statistically analyzed” or qualitative “research that emphasizes elaborate description of social or instructional settings” (Slavin, 2007, p. 8-9). The legislation previously discussed on a federal and state level allows for limited historical analysis. Thus a blend of research options were utilized to answer the questions:

1. How should the growth of music students in the state of Ohio be effectively assessed? and
2. How should the effectiveness of music teachers in the state of Ohio be determined?

One of the challenges of assessing music educators and students is the lack of consensus in assessing music students. However, there are many other content areas and teachers that do not have an established method for measuring growth. Non-tested subjects and grades include subjects with standards that cannot be adequately or completely measured with a paper-and-pencil test (e.g., art, music, industrial arts, drama, dance); subjects in lower elementary grades... students cannot be reliably tested with paper-and-
pencil or computerized tests; subjects/grades for which states have chosen not to
test because of cost and priority relative to “core” academic subjects. (Goe &
Holdheide, 2011, p. 2)

Attempting to assess music students solely through a pencil-and-paper test would
disregard the primary functions of performing, creating, and experiencing music.
Through actively listening to and creating music, students could also access an emotional
response to the art they are experiencing.

Different content areas require distinct thought processes and separate student
evaluations. Assessing music students has rarely been a one-size-fits-all solution, and
nearly impossible for music teachers in Ohio who hold a license to teach grades K-12.
Music and other fine arts share this unique characteristic, the creating and defining an
abstract concept in concrete form. While music is crucial to the holistic growth of
students, music is not always considered an “academic” subject. For instance, under the
TVAAS, a teacher’s goal is to demonstrate growth in five “academic” subjects. To
ensure that the arts remain part of the academic core, music-advocacy becomes all the
more important.

Assessing Student Growth

For teachers to establish student growth, there must be a common curriculum for
the grade/subjects taught within a school and district. Content standards should be
clearly defined, establishing what students should know at a specific point in time.
Content standards can be adapted for individual schools and districts, but the clarity of
what students should be able to do must remain. “If teachers do not have standards and a
curriculum for the grade/subject, then they must first agree on what students should know...
and be able to do ...before they can determine what different levels of performance should look like” (Goe & Holdheide, 2011, p. 17).

A value- added method is widely accepted as a means to demonstrate student growth. In music, however, there are many other skills that are not best verified through answering questions. Composing and improvising are two skills that are specifically listed within the national standards for music education (Music Educators National Conference, 1994). The recently administered NAEP test for the arts had only one question concerning composition; the test did not require students to perform in any way (Keiper, Sandene, Persky & Kuang, 2009).

Knowing various facts, performers, and symbols of music are an important facets of music, but those are only small aspects of a larger concept. Music was written with the intention of performance; it is not simply read or analyzed, but brought to life through the skills students learn from their teachers. Experiencing music as an active audience member takes a very different skill than experiencing music as an active performer.

Students must be given the opportunity to gain knowledge about music, such as listening and researching a specific topic, to experience creating music themselves, and to reflect upon their personal experience with music. Reflection requires a higher level of thinking and analysis. The reflection can come in the form of a Likert scale rubric or in free responses. There is potential for collaboration with English/Language Arts teachers to help students construct an appropriate free response that allows students to accurately reflect on their experience.

Goe and Holdheide found that it is important to “identify the specific knowledge and skills that students need to know to successfully demonstrate mastery of a particular
standard and then identify or develop tasks to serve as pretest from which progress on those standards can be determined” (2011, p. 13).

With any student evaluation, there should be a clear reflection back to the Race to the Top (federal) requirements that student learning must be rigorous, comparable across classrooms, and growth is demonstrated between two points in time (Goe & Holdheide, 2011).

Music Student Growth (Proposed)

Assessing student growth is specifically addressed as 50% of a teacher’s evaluation in Ohio under the OTES. Detailed criteria for music student growth should be developed by teachers in conjunction with LEA. The specific scoring techniques must be the same across a district, keeping with the Race to the Top requirement that growth measures are rigorous, between two points in time, and comparable across classrooms (Goe & Holdheide, 2011). Scoring rubrics should be established with administrative and collaborative approval of music teachers. Based upon existing literature and reports, a proposed rubric was designed:
There is a challenge in developing a system that measures “true performance in a way that minimizes random variation, as well as undesired and unintended consequences” (Lavy, 2007, p. 103). This can be incredibly problematic in music. Unlike other content areas, music teachers might fill classrooms of 60 students with instrument in hand or possibly teach the entire school within the year. There is little-to-no random variation for an unbiased review of music educators. In some districts, students might opt to pursue visual arts over music; in others, students are placed in music or chorus class due to student behavior issues or scheduling concerns. Without the randomization within a classroom, results might not accurately reflect the quality of teaching.
To quantify a student’s understanding of content, a variety of evaluations can take form. Tradition pencil and paper tests will not be able to cover all aspects that a music teacher might want to include in her class. Instead, portfolios and classroom artifacts can demonstrate skills that students are expected to achieve. At an age appropriate level, students can also begin a self-assessment process, evaluating their effort for a task or other subjective goals.

**Assessing Music Teacher Performance**

Measuring what music teachers specifically do in the classroom is as complex as the profession. Should an effective teacher be measured by the quality or quantity of student work produced? Should an effective teacher be measured by student behaviors or by teacher certification and professional development? In 2010, Darling-Hammond supported performance assessments for measuring what teachers actually do in the classroom. “Performance assessments...have been found to be related to later teacher effectiveness, are a much more potent tool for evaluating teachers’ competence and readiness, as well as for supporting needed changes in teacher education” (p. 5).

Many districts utilize a classroom observation as part of a teacher evaluation. “Current measures for evaluating teachers are not often linked to their capacity to teach...[evaluations] rely almost exclusively on classroom observations by principals who differentiate little among teachers and offer little useful feedback, or focus on teachers’ course-taking records and on paper-and-pencil tests of basic academic skills and subject matter knowledge that are poor predictors of later effectiveness in the classroom” (Darling-Hammond, 2010, p. 2).
Using solely subjective or objective models of evaluation poses serious concerns. For instance, if the evaluator is not effectively trained, the evaluator might not comprehend a musical activity’s purpose, potentially lowering a music teacher’s score. Likewise, an objective model such as a value-added assessment, doesn’t illustrate the process of teaching, but a final product in the form of a score. However, a combination of subjective and objective models could lead to an efficient and effective means of evaluating a teacher. “Selective emphasis can be assigned to each subjective and objective evaluation criterion, depending on subject taught or specific task, and associated with individual, team, and school performance” (Lavy, 2007, p. 101).

Much like applying educational standards for students, standards are established for good or effective teaching. The state of Ohio lists seven standards for educators (Ohio Department of Education, 2007). The National Board for Professional Teaching Standards (NBPTS) outlined eight standards specifically for music educators (2001). However, all of these standards directly correlate with Charlotte Danielson’s four domains of teaching responsibility.

Jonah Rockoff and Cecilia Speroni, economists who study educational effectiveness, found that “observation-based teacher evaluations, especially standards-based evaluation that carefully measure specific dimensions of teaching, have been found to be significantly related to student achievement gains” (Darling-Hammond, 2010, p. 7). Standards-based evaluations directly relate to the model set up by Charlotte Danielson (2007), specifically under the planning and preparation domain. Teachers can focus on the content standards for their specific lessons.
Danielson’s four domains share numerous qualities with the state of Ohio standards for educators, NBPTS, and OTEA. One could apply the five points of effective teaching (Little, Goe, Bell, 2009) to the Danielson domains— from maintaining high expectations measured by value-added means (Domain 1 and 3); contributing to a positive environment academically and socially (Domain 2); using a variety of assessment and learning resources for students and adapting lessons as needed (Domains 1 and 3); developing environments that respect diversity and “civic-mindedness” (Domain 2); and collaborating with other educational professionals (Domain 4) (Danielson, 2007). Based upon these trends, the following chart was drawn:
FIGURE 2: Domain Overlaps

<table>
<thead>
<tr>
<th>Danielson Domain 1: Planning and Preparation</th>
<th>Danielson Domain 2: The Classroom Environment</th>
<th>Danielson Domain 3: Instruction</th>
<th>Danielson Domain 4: Professional Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ohio Standard 1: Students</td>
<td>Ohio Standard 1: Students</td>
<td>Ohio Standard 2: Content</td>
<td>Ohio Standard 6: Collaboration and Communication</td>
</tr>
<tr>
<td>Ohio Standard 1: Assessment</td>
<td></td>
<td>Ohio Standard 4: Instruction</td>
<td></td>
</tr>
<tr>
<td>OTES: Communication and Professionalism</td>
<td></td>
<td></td>
<td>OTES: Communication and Professionalism</td>
</tr>
</tbody>
</table>

The Ohio Department of Education established the Ohio Teaching Evaluation System (OTES) which divides a teacher’s evaluation into two segments; half of a teacher’s evaluation is based on teacher performance, half based on student growth measures (2011). The following is a proposed rubric for the teacher performance aspect of OTES. OTES established the four point system of “Accomplished” (the highest designee), “Proficient”, “Developing”, and “Ineffective” for the overall evaluation of teachers.
FIGURE 3: Assessment for Music Teachers in the State of Ohio (Proposed)

The average of the rating can establish an overall rating for a teacher. According to the OTES, teachers will “participate in a minimum of two formal observations” (Ohio, Ohio Department of Education, 2011, p. 14). Combined with student growth, teachers can earn the designee of Accomplished, Proficient, Developing, or Ineffective. Specific details of reevaluations processes after implementing an improvement plan, if necessary, need to be part of an established agreement between teachers/LEA and district administration and follow Ohio Department of Education policies.
CHAPTER 4
DISCUSSION

Connecting teacher performance to student growth would seem to satisfy policy makers and community members, offering a “business metaphor of contemporary accountability discourse that views test scores as accurate measures of educational „value”” (Kupermintz, 2003, p. 287). Unlike some corporate entities, however, the product that educators offer cannot be measured by profit analysis. Music educators, in particular, are challenged to place a value and scale on abstract concepts. Administrators and music educators must answer the following questions:

1) How should the growth of music students in the state of Ohio be effectively assessed? And

2) How should the effectiveness of teachers in the state of Ohio be determined?

There are a variety of methods and instruments used to determine teacher effectiveness. But what is the true purpose of a teacher evaluation? Depending upon how the data are gathered and reported, a variety of conclusions could be reached. If effectiveness is the purpose, no matter how “effectiveness” is defined and whether formative or summative, the ultimate goal of teacher evaluation should be to improve instruction. There is no single measure that can be utilized to establish a comprehensive view of teacher effectiveness and student growth. Teacher effectiveness and student growth are two sides of the same coin. Students need the guidance of effective teachers
to improve and grow; teachers need students to provide feedback on whether pedagogical techniques are appropriate for a specific group of students (Kupermintz, 2011). While the fate of student and teacher might be linked, student growth and teacher effectiveness are two separate evaluations.

Nonetheless teacher evaluations can offer vital information for teachers, administrators, students, and parents. It is important to determine the desired outcome of the evaluation before an evaluation tool is selected (Little, Goe, & Bell, 2009).

Any teacher evaluation tool is going to cost time and resources to ensure that the evaluations are carried out in a timely and efficient manner. Evaluators must be properly trained. Creating a new evaluating system for a school or district results in additional costs upfront. But can, or should, a price be placed on an effective teacher?

From a purely statistical perspective, music teachers are at a direct disadvantage from their peers. Often there is one music teacher per school or even the whole district. Depending on the district curriculum, music could be compulsory at all ages or could be an elective for students. Non-random assignment or student bias can certainly skew statistical results (Kupermintz, 2009; Lavy, 2007). A music teacher does not always receive a balanced classroom of students; ensemble directors are asked to differentiate instruction for a variety of group sizes- from a chorus of 12 to a marching band of 150.

There is a great concern that if an evaluation model is created without certain standards in place, teachers and evaluators could be prone to “game” the system for a specific, desired result (Lavy, 2007). Teachers, in order to increase pay or save their jobs, might find themselves teaching to the test, simplifying questions or content, or limiting student exposure to valuable, yet non-assessed material. Evaluators could
effortlessly play a bias and intentionally look for specific data to support the bias. Either way, both teacher and evaluator need to reflect upon the overarching concern—*Is this best for my students?* Lavy, in 2007, warns of a greater focus on specific skills linked to testing scores, thus, “[Sacrificing] the nurturing of curiosity and creative thinking to teaching the skills tested on standardized exams... teaching to the test” (p. 92). Music teachers might be concerned that such a statistic or judgment is passed on what is a subjective or emotion filled opportunity.

Music teachers must be very aware of teaching to the whole child and not just to one aspect of music that data is required. Students deserve to have a well-rounded education that balances skill with knowledge and offers time to reflect upon the impact of music within their lives. “When one dimension of output is easily measured but another is not, teachers may dedicate their efforts to maximizing the measurable at the expense of the unmeasured dimension” (Lavy, 2007, p. 92). All dimensions of teaching music must be assessed.

Rewards of teaching are often monetary, but can also come in the form of reduced teaching load, public acknowledgement (such as an award of excellence) or some system of promotion. Rewards can occur only once or rewards can be ongoing. But, “If wages are based on student performance, they provide teachers or schools with powerful signals about what is valued and what is not” (Lavy, 2007, p. 90). All content areas, those already tested and non-tested, are important for whole growth and development of the student.

Another concern is that the value-added performance pay measures are rewarded on a short-term basis. It is difficult to track student performance throughout their entire
K-12 instruction and beyond. It is equally challenging to evaluate if skills acquired for proficiency in elementary school translated to proficiency in the secondary education and into the labor market. Again, we must ask ourselves what are our overarching goals? If the goal of education is to successfully prepare students for the labor market after graduation, then do these short-term goals for teachers have any effect on student success later in life?

How performance pay initiatives will be funded has yet to be determined. Lavy brings up numerous concerns of teacher performance pay:

If wages are based not only on the individual benefits of schooling to students, but on the benefits to society as a whole, teachers or schools could take into account the social return to education when making choices about their work... Teachers may focus disproportionate attention on those students who are most likely to improve their test scores or to cross a designated threshold. (2007, p. 90)

In order for the evaluation process to be effective, efficient, and just, it will take additional funds. Cash strapped districts could be forced to set aside large amounts of funding for teacher evaluations, instead of new resources or programs for the classroom for students.

For an effective performance pay system to work, “all teachers should be eligible for the incentive offered... however, only a subset of teachers should be rewarded in practice” (Lavy, 2007, p. 102). This ensures that only the highest quality of teaching is recognized. Teachers not at the highest level of instruction should not be awarded as such.
It should be noted that no teacher has an easier job than the next. All content areas are equally challenging, especially in this age of accountability. With appropriate time, effort, and planning of lessons, teaching is equally challenging, and rewarding. There is limited information available with regards to music teacher assessments. The proposed rubrics for student growth and music teacher assessments could greatly assist the field of music education, but these tools need an opportunity to be piloted in a variety of music settings. Additional research and studies are needed in this area.

“No person should be evaluated for high-stakes decisions based on statistical assumptions rather than on actual information” (Darling-Hammond as cited in Kupermintz, 2003, p. 291). Collecting data and a variety of evidence is critical when assessing teachers. Teachers can also value the opportunity for constructive commentary of their teaching. An effective evaluation is unbiased and offers information of the positive areas in which teachers excel and honest assessment of areas that teachers can improve. School administrators should keep open communication to teachers as to the school and district expectations for all teachers, as well as offer opportunities for teachers to grow and learn as needed.
CHAPTER 5
EDUCATIONAL IMPLICATIONS

Utilizing a rubric approved by districts, administrators, and teachers offers all parties involved an opportunity for input as the exact measures required for music student evaluations. Each school district is different, based on the community it serves. Music teachers are very aware of this fact. There are only a handful of teachers who see the vast number of students within an academic year. Establishing a specific rubric and measures of growth across the state is unrealistic. Individual school districts utilize various resources, in both human and capital form. Music, as one of the fine arts, qualifies under NCLB legislation as an “integral part of elementary school and secondary school curriculum” (No Child Left Behind Act of 2001, 2002). For example, the critical thinking skills developed in music classes are a 21st Century skill, from a skill set required of other content areas.

The opportunity to develop specific rubrics by district and grade level allows teachers to have ownership of the process. The overarching guiding process of creating any assessment tool should be State and National standards. And like State and National standards, teachers and administrators need to review the student and teacher assessment processes regularly, and be willing to make any and all changes necessary for an improved assessment. An unwritten skill of education is the need for flexibility.
Teachers need to be reminded that these assessment tools are not intended to rate or rank, but to highlight areas in which teachers and student are excelling, and areas for improvement (Ohio Department of Education, 2011). An assessment is, after all, “any systematic method of obtaining information about knowledge and skills and using it to draw inferences about people or programs” (Tung, 2010, p. 50). To be a better teacher should be the goal of all teachers. Assessing how to be a better teacher will require constant reflection and revision.

Whether or not political interaction can or should “reform” education is a matter of philosophical debate. However, at the core of some of the legislation asks teachers, administrators, and parents to quantify and qualify the characteristics of effective teaching. Music teachers can and should be held accountable for the work done in and out of the classroom. Since one of the many tasks charged to a teacher is ensuring student growth, requiring an analysis of data is quite reasonable. There are many other characteristics, far more subjective than analyzing data, which can influence a teacher’s effectiveness. However in a social climate that expects a clear, bottom line results, music teachers must compile and analyze data with their other academic cohorts.

Music teachers need to participate fully in the assessment discussions. Currently as a non-tested content area, there is a level of flexibility, due to the lack of state requested information. As districts develop a common curriculum, assessments can be created and an appropriate growth measure established. Teacher accountability is here to stay for the time being. Lawmakers, parents, and administrators wish to work in concert with teachers to create a positive, productive environment for students. Likewise, music teachers need to remind themselves to set aside personal philosophies and put the
student’s best interest first, a step forward in being a more effective teacher. Music
teachers must take an honest look at their classrooms, students, and practices and answer
the question: *Am I an effective music teacher?*
REFERENCES


Ohio Office of Budget and Management. (2011). *State of Ohio the executive budget*


