

MÜLLER, STEPHEN W. Ph.D. A Descriptive Analysis of High School Music Teacher Evaluation in the Commonwealth of Virginia. (2022)
Directed by Dr. Brett Nolker. 140 pp.

The purpose of this descriptive analysis study was to investigate high school music teachers' levels of satisfaction with the teacher evaluation practices used by administrators and music supervisors in the Commonwealth of Virginia. Virginia high school music teachers ($n = 76$) were surveyed to collect data used to determine their levels of satisfaction with their evaluation and what, if any, factors are related to their levels of satisfaction. The satisfaction level was measured using a researcher-constructed *Teacher Evaluation Satisfaction Survey (TESS)*.

Participants responded to prompts on satisfaction with the process, personnel, and product of evaluation. Slightly over half the participants reported satisfaction with the process ($M = 3.5$), personnel ($M = 3.4$), and product ($M = 3.1$) of their evaluation. Linear regression analysis showed a significant difference in levels of satisfaction with the process of their evaluation ($F(2,52) = 60.82, p < .001, R^2 = .70$) by participants who believed that their evaluation highlighted teacher practices and that their evaluation criteria were appropriate. Similarly, participants who indicated a level of trust with their evaluator who had experience with non-tested grades and subjects showed a significant difference in levels of satisfaction with the evaluation personnel ($F(2,51) = 53.17, p < .001, R^2 = .68$). Participants who believed that their evaluation accurately summarized their performance and led to improvements in student learning showed a significant difference in levels of satisfaction with the product of their evaluation ($F(2,51) = 37.05, p < .001, R^2 = .59$).

A DESCRIPTIVE ANALYSIS OF HIGH SCHOOL MUSIC TEACHER EVALUATION
IN THE COMMONWEALTH OF VIRGINIA

by

Stephen W. Müller

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

Greensboro

2022

Approved by

Dr. Brett Nolker
Committee Chair

© 2022 Stephen W. Müller

DEDICATION

I would like to dedicate this work to my family, friends, and colleagues who have accompanied me on this life-journey which often includes academic projects, degrees, and dissertations. We have all been a part of this music education laboratory as we pursue our professional and academic endeavors, and I am a better person, husband, father, and educator because of YOU!

To my wife, Valerie, I would like to offer my deepest gratitude for being my constant companion on this journey- WOW, what an adventure! You have earned this degree as much as I have, and I would not be where I am today if it were not for you. I love you!

To my children, Abigail, David, Caleb, and Kaylee, thanks for making me a dad! My job, career, and degrees are all great, but being your dad is one of my proudest accomplishments. You have inspired me to do better and be better. I love you!

APPROVAL PAGE

This dissertation written by Stephen W. Müller has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair

Dr. Brett Nolker

Committee Members

Dr. Patricia Sink

Dr. Tami J. Draves

Dr. Jennifer Walter

November 3, 2022

Date of Acceptance by Committee

October 20, 2022

Date of Final Oral Examination

ACKNOWLEDGEMENTS

It might seem redundant to acknowledge the faculty and staff of the UNCG School of Music, but I feel I must. As a professional educator, I have been impressed from day one with the level of expertise and professionalism found in every aspect of this program. Attention to detail in the small things and excellence in the biggest things is hard to achieve. I am honored to call these nationally known and frequently published authors my professors, colleagues, and friends.

Specifically, Dr. Patricia Sink and then Dr. Brett Nolker, my Dissertations Chair(s) (yes, it took that long!). Thank you for your great professionalism mixed with a personal touch. Thank you for taking the time to get to know me and my context and for your willingness to work with me from start to finish on this project.

I would also be remiss if I did not mention the faculty, staff, and students of the School of Music at Liberty University, I am honored and humbled to serve as your Dean. Thank you for letting me “practice” my craft and implement so many of the things I have learned in this degree in my role at LU- GO FLAMES!

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xi
CHAPTER I: INTRODUCTION.....	1
Purpose of the Study	1
Research Questions	2
Background of the Study.....	2
History	3
Law	4
Implications	4
Advocacy	7
Definition of Terms	8
CHAPTER II: REVIEW OF LITERATURE	11
Introduction	11
Teacher Evaluation.....	12
Formative and Summative Teacher Evaluation	18
High-Stakes Teacher Evaluation	23
Non-tested Grades and Subjects	27
Music Teacher Evaluation.....	30
Use of Value-added Measures	37
Summary	41
CHAPTER III: METHOD	44
Restatement of the Purpose	44
Participants	44
Participant Selection	45
Participant Demographics	45
The Survey Instrument	45
Survey Pilot	51
Survey Distribution	52

Data Analyses.....	52
Part 1.....	52
Part 2.....	52
Part 3.....	53
Summary	54
CHAPTER IV: RESULTS.....	55
Descriptive Analyses of Data.....	56
Demographics.....	56
Age.....	56
Sex	56
Race/Ethnicity.....	56
Education Level	56
Years of Service.....	56
Teacher Evaluation Process.....	60
Teacher Evaluation Personnel	64
Teacher Evaluation Product	66
Research Questions	71
Research Question 1	71
Teacher Evaluation Process	72
Teacher Evaluation Personnel.....	72
Teacher Evaluation Product.....	73
Research Question 2	74
Demographic Effect.....	74
Teacher Evaluation Process	76
Teacher Evaluation Personnel.....	78
Teacher Evaluation Product.....	80
Summary of Results	82
Research Question 1	82
Research Question 2	83
CHAPTER V: DISCUSSION.....	85
Research Question 1	86

Research Question 2.....	88
Teacher Evaluation Process.....	89
Teacher Evaluation Personnel.....	89
Teacher Evaluation Product.....	90
In Their Own Words.....	90
Possible Limitations.....	91
Implications.....	92
Suggestions for Future Research.....	95
Recommendations.....	96
Conclusion.....	97
REFERENCES.....	99
APPENDIX A: NafME AND MENC POSITION STATEMENTS.....	109
APPENDIX B: TEACHER EVALUATION SATISFACTION SURVEY.....	113
APPENDIX C: OPEN-ENDED RESPONSES: IN THEIR OWN WORDS.....	119
APPENDIX D: LINEAR REGRESSION ANALYSIS ON TEACHER EVALUATION PROCESS.....	132

LIST OF TABLES

Table 1. <i>TESS</i> Part 1	47
Table 2. <i>TESS</i> Part 2	50
Table 3. <i>TESS</i> Part 3	51
Table 4. Likert-type question responses	53
Table 5. Demographic Information.....	57
Table 6. Teaching Context.....	58
Table 7. Teaching Context – subject(s) taught	59
Table 8. Teaching context – personnel and frequency	60
Table 9. Teacher Evaluation Process	61
Table 10. Factors, practices, or characteristic that have a positive impact on your levels of satisfaction with music teacher evaluation	63
Table 11. Factors, practices, or characteristic that have a negative impact on your levels of satisfaction with music teacher evaluation	64
Table 12. Teacher Evaluation Personnel	65
Table 13. "Good" evaluator	66
Table 14. Summary Statistics Table for Interval and Ratio Variables	68
Table 15. Advice for new teachers	69
Table 16. Advice for evaluators.....	70
Table 17. Reported levels of satisfaction with evaluation	71
Table 18. Reliability Table for Satisfaction Composite.....	74
Table 19. Results of Kruskal-Wallis analyses	74
Table 20. Results of Kruskal-Wallis analyses	75
Table 21. Results of Kruskal-Wallis analyses	75
Table 22. Variance Inflation Factors for Q13.4 and Q13.8	77

Table 23. Results for Linear Regression with Q13.4 and Q13.8 predicting Q13.9	78
Table 24. Variance Inflation Factors for Q15.2 and Q15.5	79
Table 25. Results for Linear Regression with Q15.2 and Q15.5 predicting Q15.6.....	80
Table 26. Variance Inflation Factors for Q18.1 and Q18.6	81
Table 27. Results for Linear Regression with Q18.1 and Q18.6 predicting Q18.7.....	82
Table 28. Reported levels of satisfaction with evaluation	86
Table 29. <i>TESS</i> Results to National HS Teachers.....	95
Table D30. Variance Inflation Factors for Q13.1, Q13.2, Q13.3, Q13.4, Q13.5, Q13.6, Q13.7, and Q13.8	133
Table D31. Results for Linear Regression with Q13.1, Q13.2, Q13.3, Q13.4, Q13.5, Q13.6, Q13.7, and Q13.8 predicting Q13.9.....	134
Table D32. Variance Inflation Factors for Q15.1, Q15.2, Q15.3, Q15.4, and Q15.5.....	136
Table 33. Results for Linear Regression with Q15.1, Q15.2, Q15.3, Q15.4, and Q15.5 predicting Q15.6	137
Table D34. Variance Inflation Factors for Q18.1, Q18.2, Q18.3, Q18.4, Q18.5, and Q18.6.....	139
Table D35. Results for Linear Regression with Q18.1, Q18.2, Q18.3, Q18.4, Q18.5, and Q18.6 predicting Q18.7.....	140

LIST OF FIGURES

Figure 1. Q-Q scatterplot for normality of the residuals for the regression model.....	77
Figure 2. Q-Q scatterplot for normality of the residuals for the regression model.....	79
Figure 3. Q-Q scatterplot for normality of the residuals for the regression model.....	81
Figure D4. Q-Q scatterplot for normality of the residuals for the regression model.....	132
Figure D5. Q-Q scatterplot for normality of the residuals for the regression model.....	135
Figure D6. Q-Q scatterplot for normality of the residuals for the regression model.....	138

CHAPTER I: INTRODUCTION

Many industries and businesses utilize quality control measures to test or evaluate their products or services. These measures indicate effectiveness or show areas in need of improvement. For professional educators, formal, systematic teacher evaluation has become the primary method of assessing teacher effectiveness. Accurate and effective teacher evaluation is an essential component of job satisfaction which has been linked in research to longevity. Longevity for effective teachers and a process to improve or remove ineffective teachers positively affect student learning outcomes (Adnot, Dee, Katz, & Wyckoff, 2017). This is why it matters; teachers are the most important factor of all the elements of a classroom that impact student success (Opper, 2019).

Concerns for accurate music teacher evaluation have been expressed in publications by various professional organizations, and to a lesser extent, in the research literature. In a position statement by the National Association for Music Education (NAfME), essential components for accurate music teacher evaluation included measuring student growth within the music teaching content area and observations by personnel knowledgeable in music. Since the federal mandate (ESSA, 2015) to include student growth measures in the evaluation of teachers, various means of measuring student growth in music have been discussed including use of portfolios and contest ratings (Hash, 2013; Parkes et al., 2015; Potter, 2021). The current study will add to the body of literature on music teacher evaluation by investigating music teachers' levels of satisfaction with how they are evaluated.

Purpose of the Study

The purpose of this study was to investigate high school music teachers' levels of satisfaction with the evaluative practices used by administrators and music supervisors in the

Commonwealth of Virginia. Virginia high school music teachers were surveyed to collect data to determine their levels of satisfaction with their evaluation and what, if any, factors related to their levels of satisfaction. The satisfaction levels were measured using a researcher-constructed *Teacher Evaluation Satisfaction Survey (TESS)* completed by high school music teachers in Virginia. The data collected were used to answer the research questions presented in this study.

Research Questions

1. What are the levels of satisfaction, among high school music teachers, with contemporary evaluative practices?
2. Are there specific factors that relate to the levels of satisfaction with teacher evaluation?

Background of the Study

The problem addressed in this study was the potential concerns raised in evaluating music teachers with the same tools used for evaluating classroom teachers in other subject areas. In most teacher evaluation processes, one evaluation rubric, with one set of measures and corresponding expected outcomes is used for all teachers, regardless of content area. Salvador and Krum (2019) shared the concern that best practices in some music classes “are not necessarily the same as the teacher behaviors described by indicators on teacher evaluation rubrics” (p. 146). The problem created by a lack of discipline-specific evaluation rubrics is compounded if the evaluating personnel lack content knowledge or experience in evaluating classes outside their expertise. Salvador and Krum pointed out the unrealistic expectation placed on an evaluator to know the philosophies and practices in an area outside their experience. They warned of the increased likelihood of failing to recognize teaching and learning interactions in the music context, the most critical component of observation and evaluation.

History

The importance of teacher evaluation as a means of assessing and preserving quality education can be traced back to the early 1700s (Marzano et al., 2011). Early American town officials or clergy served as supervisors in hiring, firing, and evaluating teachers. Marzano et al. noted a change in supervision of education to a more formalized and professional process as public school size and availability grew. A general connection or correlation between teacher practices and student outcomes can be seen in the adage “So goes the teacher, so goes the school” found in the 1845 *Annual Report of the Superintendent of Common Schools of the State of New York* (Blumberg, 1985, p. 58).

Music education was added to the curriculum in Boston in 1838 when Lowell Mason was hired to teach music (Kelly, 2016). Kelly (2016) cited four lessons to be learned from Lowell Mason: (1) everyone has music ability, (2) the power of the individual teacher to make a difference, (3) the ability of community action to keep music in schools, (4) schools and communities must work together for music to be at its greatest influence (p. 49). Most relative to the current study was item two, the power of the individual teacher to make a difference. Music education in Boston served as a model of music in the schools. “The Boston music program included supervision of classroom teachers, teacher-training classes, textbook publication, and performances” (Howe, 1992, p. 316).

In 1910, the first meeting of the Music Supervisors National Conference¹ was held in Iowa. Among the purposes of their initial meeting was concern over a lack of trained music teachers in the schools (Kelly, 2016). Although noted by Lowell, the link between individual

¹ The Music Supervisors National Conference became the Music Educators National Conference in 1934 and then the National Association for Music Education (NAfME) in 2011 (<https://nafme.org/about/>).

teaching effectiveness and student learning outcomes was more prominently featured much later. “The conclusion that individual teachers can have a profound influence on student learning even in schools that are relatively ineffective, was first noticed in the 1970s when we began to examine effective teaching practices” (Marzano et al., 2001, p. 3).

Law

Federal education laws and initiatives, including No Child Left Behind (NCLB, 2002-2015), Race to the Top (RTTT, 2009), and Every Student Succeeds Act (ESSA, 2015) mandated student growth measures be included in teacher evaluation. Accountability was placed with individual states to determine how student growth measures are included. The ESSA mandates required that music teachers be evaluated based on student growth measures outside the content of the music classrooms and music teacher practices.

Implications

With the inclusion of student growth measures in teacher evaluation, the process became more complicated for all teachers. Two concerns rose to the surface: (1) by which student growth measures should teachers of non-tested grades and subjects (NTGS) be evaluated? and, (2) how much student growth can be accurately related to teacher practices?

First, by which student growth measures should teachers of non-tested grades and subjects (NTGS) be evaluated? The Arts (drama, dance, music, and theater), physical education, and foreign languages are among these non-tested subjects. To comply with early mandates, some districts divided teachers from the non-tested subjects into groups and added test results from available standardized tests in the evaluation of these teachers. The result was a teacher evaluation system based partially on student growth outside the content area of these teachers.

This concern was documented as early as 1990 when Taebel (1990) cited concerns over the applicability of teacher competencies to the music teaching context and questions about evaluator qualifications if they were not trained in music. Taebel concluded that the explanation for consistently low-scoring performance by music teachers could be attributed to one of two reasons: (a) music teachers lack the professional competencies of their non-music teaching colleagues, or (b) the specific teaching behaviors required for success in music education are not recognized by the evaluation system. This concern grew in intensity, and several decades later, a similar study by Shuler (2012) echoed even more strongly, “Evaluating teachers of music based on test scores in subjects they do not teach is not only unfair; it is absurd. Evaluating them based on schoolwide scores that include students they do not teach compounds the injustice” (p. 9).

The second concern raised was how much student growth can be accurately related to teacher practices? Numerous researchers have emphasized that teachers are the most influential factor in student learning within the school context (Hewitt & Amrein-Beardsley, 2016; Marzano et al., 2001; Robertson-Kraft & Zhang, 2018). However, the relationship between teacher effectiveness and student growth often indicates an overly simplistic understanding of this relationship which can lead to the conclusion that the way to increase student growth is simply to improve teacher quality. This error in logic does not consider the myriad of confounding factors impacting student success that stand between teacher practices and student growth. These factors include socioeconomic status, peer pressure, family background, limited educational opportunities, and more (Marshall, 2013, Salvador and Krum, 2019).

An incomplete understanding of a complex relationship between student outcomes and teacher quality resulted in legislation to address the problem by implementing well-intentioned, one-size-fits-all reforms. Differences in state educational systems, socioeconomics,

urban/suburban/rural location, school size, and other demographic characteristics create a level of complexity almost guaranteeing ineffectiveness, if not failure, of these teachers, administrators, and practices. Chester Finn (2015), educational policy analyst, and former United States Assistant Secretary of Education, wrote about a new complacency growing among education policymakers and reformers. The basis of his critique was that the creation and implementation of new laws or programs was not the end of the process, but rather, the beginning and needed to be followed up with evaluation and modification as needed for maximum effectiveness (Finn, 2015).

Further evidence that increased emphasis on teacher evaluation did not guarantee a direct and positive correlation to increased teacher effectiveness was illustrated by the *widget effect*. There was often an underlying assumption that a “good” or “great” school must by default be staffed by “good” or “great” teachers. A seminal report on teacher evaluation as a measure of teacher effectiveness (Weisberg et al., 2009) called out the failures of teacher evaluation systems to accurately identify effective and ineffective teachers.

The failure of evaluation systems to provide accurate and credible information about individual teachers' instructional performance sustains and reinforces a phenomenon that we call the Widget Effect. The Widget Effect describes the tendency of school districts to assume effectiveness is the same from teacher to teacher. This fallacy fosters an environment in which teachers cease to be understood as individual professionals, but rather as interchangeable parts. In its denial of individual strengths and weaknesses, it is disrespectful to teachers; in its indifference to instructional effectiveness, it gambles with the lives of students. (Weisberg et al., 2009, p. 32)

The Weisberg et al. study highlighted the challenges facing teachers and teacher evaluation in a broad and general sense. Most teacher evaluation systems were designed for a specific learning context, and therefore any learning environment that differs from the assumed context may suffer as the result. Music classrooms and teaching contexts add additional layers of complexity that may be hard to evaluate with commonly used evaluation systems.

Administrators or other evaluators with limited experience or content knowledge in a subject may struggle to identify professional competencies within the context of that subject and classroom.

Advocacy

A 1988 policy statement on music teacher evaluation from the Music Educators National Conference (MENC) affirmed the role and importance of teacher evaluation but called for modification of evaluation procedures to “adequately evaluate the performance competency of music teachers” (MENC, 1988). The leaders of MENC also advocated for using special assessment instruments for music teacher evaluation by evaluators knowledgeable of music teaching and learning. In the current position statement, the National Association for Music Education (NAfME, former MENC) cautions against the use of student scores not related to music content or music teaching practices to evaluate music teaching. The position statement outlines criteria for measures of student achievement and teaching characteristics to validly evaluate music teachers (Appendix A).

In a position statement on “Assessment in Music Education,” the National Association for Music Education (NAfME) has identified a potential challenge facing music teachers in “an environment where dependence on large-scale, high stakes testing of students in a narrow range of programs (reading, math, and science) threatens to limit learning opportunities to those tested

subject areas” (NAfME, 2009, para. 2). The position statement further calls for student growth in the music classroom to be measured relevant to the subject(s) taught in the music classroom.

...some form of regular assessment of music programs should be adopted. The assessment should measure student learning across a range of standards representative of quality, balanced music curriculum, including not only responding to music but also creating and performing music. This assessment should serve the goal of educational accountability by providing data that can be included in the school- or district-level “report card” disseminated to the public as required by law. (para. 7)

Definition of Terms

In any trade or profession, words or phrases commonly used are often referred to by their acronym. Other terms have discipline-specific meanings or connotations. The following list contains commonly used words, phrases, and titles with their definition or a brief explanation.

- *teacher evaluation* - a systematic, ongoing process used to assess teachers’ competence, performance, and effectiveness in the classroom (Jenkins, 2018)
- *non-tested grades and subjects (NTGS)* – grade levels and subject areas not assessed by standardized tests
- *high-stakes teacher evaluation (HSTE)* – a teacher evaluation system that allows districts to identify the effectiveness of teachers in order to inform dismissal and layoff decisions, removal and grant of tenure, promotion, and bonuses (Lavigne, 2014)
- *value-added measures (VAM)* - a sophisticated statistical model designed to translate student test scores into teacher effectiveness estimates (Haertel, 2013)

- *student learning objectives (SLOs)* – teacher-developed measures of student performance (Steinberg & Donaldson, 2016)
- *No Child Left Behind (NCLB)* – a 2002 update of the Elementary and Secondary Education Act intended to close the achievement gap by holding states and individual students accountable for student learning (Lavigne, 2014)
- *Race To The Top (RTTT)* – a \$4 billion dollar program that awards states with coherent and rigorous plans to prepare students for success in college and the workplace and to compete in the global economy; build data systems that measure student growth and success and inform instruction; recruit, develop, reward, and retain effective teachers and principals; and finally, turn around the lowest-achieving schools (U.S. Department of Education, 2012)
- *Every Student Succeeds Act (ESSA)* – the 2015 replacement of NCLB gives states and districts the opportunity to move beyond NCLB’s reliance on narrow measures of school success so that they can reimagine and redefine accountability for their schools in a more holistic way that supports a high-quality education and equal opportunity for all students, and ensures they are learning to college- and career-ready expectations (U.S. Department of Education, 2016)
- *highly qualified teachers* – a key component of NCLB, a teacher who has earned a bachelor’s degree, full state certification, and demonstrated competency as defined by the state in each core academic subject he or she teaches (U.S. Department of Education, 2004)
- *well-rounded education* – consists of courses, activities, and programming in subjects such as English, reading or language arts, writing, science, technology, engineering,

mathematics, foreign languages, civics and government, economics, arts, history, geography, computer science, music, career and technical education, health, physical education, and any other subject with the purpose of providing all students access to an enriched curriculum (U.S. Department of Education, 2016)

CHAPTER II: REVIEW OF LITERATURE

Introduction

In this review of the literature, topics related to teacher evaluation were examined and more specifically, the purposes of teacher evaluation, high-stakes teacher evaluation, non-tested grades and subjects, music teacher evaluation, and the use of value-added measures. While investigating these topics, the researcher found several related state-specific studies which were reviewed to provide a deeper understanding of the impact teacher evaluation is having across the country. Many of the studies reviewed included the use of surveys to collect data regarding the attitudes and opinions participants had on the topics of teacher evaluation, job satisfaction, and other related topics.

The body of research on teaching evaluation practices showed a flurry of activity following implementation of each of the major federal education laws and incentives enacted over the past two decades. No Child Left Behind (NCLB 2002-2015) emphasized standardized tests and introduced the idea of “highly qualified” teachers as determined by a greater reliance on the results of standardized tests. Race to the Top (2009) was a \$4.35 billion federal grant offered by the U.S. Department of Education designed to incentivize states to implement innovation and reform. Relevant to this study was its emphasis on performance-based teacher evaluation (previously discussed as “high-stakes teacher evaluation”) and the use of student growth measures as a major component of determining teacher effectiveness via teacher evaluation. The use of these student growth measures is concerning to teachers of tested subjects and teachers of non-tested grades and subjects (NTGS) alike.

One effect of the Every Student Succeeds Act (ESSA, 2015) was an apparent balance between federal mandates and local implementation and oversight of teacher evaluation. A shift

in language and emphasis from “highly qualified” teachers to “well-rounded education” was notable, as well as a broadening of “core subject” offerings. In a special report published by the Education Commission of the States entitled “ESSA’s Well-Rounded Education,” this distinction is further explained. Where NCLB overemphasized the use of English and math standardized tests to measure student growth, ESSA sought a more balanced approach by incorporating the arts, humanities, and sciences (Jones & Workman, 2016). The ESSA also gave states the responsibility and flexibility to determine how teacher evaluations should be conducted. In many states (discussed later in this chapter) this led to a decrease in the weight of standardized tests and the inclusion of alternative means of evaluating student growth.

Teacher Evaluation

Jenkins (2018) defined teacher evaluation as “a systematic, ongoing process used to assess teachers’ competence, performance, and effectiveness in the classroom” (p. 1658). He outlined the general process of evaluation, including a pre-observation meeting where the time, place, and framework of the observation are discussed. Jenkins described this initial meeting in which the evaluator can communicate expectations and/or the teacher can describe the context or content of the upcoming observation. Discussion of any self-evaluation can also be held in the pre-observation meeting. The next step is classroom observation where the evaluator gathers data on the teacher’s execution of the lesson plan, classroom management, teaching style, and other aspects of professional competency and teacher effectiveness. The final part of an evaluation is the post-observation meeting where the evaluator communicates the results of the observation and any professional development actions needed.

In the white paper, “Teacher Assessment and Evaluation” published by the National Education Association (2010), the NEA provided a summary of concerns with teacher evaluation

and a framework for an improved process designed to enhance both teacher effectiveness and student outcomes (see Figure 1). The article cited the use of value-added measures² (VAMs), a sophisticated statistical model designed to translate student test scores into teacher effectiveness estimates (Haertel, 2013). Some evaluation systems have overemphasized the summative purposes – evaluate, rank, reward, and punish. The National Education Association offered a scathing review of VAMs:

Over the past several years, numerous studies have concluded that value-added methodology is neither fair enough, nor reliable enough, nor valid enough to be used as a basis for high-stakes decisions about teachers. It is difficult, if not impossible, to find any education researchers who support the use of value-added methodologies for high-stakes decisions about teachers, other than several researchers who have developed and promoted the use of value-added methodologies. (NEA, 2010, p. 8)

As outlined in this section, the number of factors impacting student learning inside and outside the classroom have created great difficulty in measuring student growth accurately. In place of an overreliance on a single measure of student growth found in many VAMs, the NEA recommended the use of multiple measures of student growth including:

- Local and district-wide achievement test results
- Student work that shows evidence of student growth
- Subject matter assessments
- Students’ oral and written presentations
- Learning goals developed by the teacher or principal

² VAMs are sometimes referred to as value-added measures, methods, or methodologies.

- Project-based inquiry activities
- Teacher-generated information about student growth and goals
- Formative and summative student assessments
- Evaluations of effective engagement, critical thinking, self-efficacy or a combination thereof (NEA, 2010, p. 9)

The underlying belief was that great teachers provide great teaching which produces great students. While this formula did not consider the myriad of factors affecting learning, it was the foundation upon which teacher evaluations were built. Teachers were identified as the most important school-related factor to student success (Oppen, 2019). Marzano et al. (2001) came to the same conclusion that teachers “can have a profound influence on student learning even in schools that are ineffective...” (Marzano et al., 2001, p. 3).

The impact effective teachers have on student learning can counteract many environmental factors outside the scope of the classroom, including social class, innate intelligence, family background, community dynamics, negative peer pressure, racism, and discrimination (Marshall, 2013). Marshall cited no less than ten studies supporting the contention that, “the quality of instruction is the single most important factor in student achievement” (p. 1). His book, *Rethinking Teacher Supervision and Evaluation*, builds a case for improving student learning through the partnership between principals and teachers to use evaluation as a tool for these five purposes:

- Appraisal – getting an accurate sense of the quality of instruction
- Affirmation – retaining and further developing highly effective teachers
- Improvement – coaching and supporting developing teachers
- Housecleaning – dismissing teachers who are not effective, even after coaching

- Quality assurance – holding a high standard and honestly reporting to stakeholders (Marshall, 2013, pp. 21-22)

The tools used with the system of teacher evaluation can vary. Among the more popularly used tools are Danielson’s Framework for Teaching (FFT) and Marzano’s Causal Teacher Evaluation Model. A brief overview reveals that Danielson’s FFT utilizes a 4-point scale to assess four domains – Planning and Preparation, Learning Environments, Learning Experiences, and Principled Teaching with a focus on teacher effectiveness. Marzano’s model consists of a 5-point scale assessing four domains – Classroom Strategies and Behaviors, Planning and Preparing, Reflecting on Teaching, and Collegiality and Professionalism with a focus on teacher development.

Nielsen (2014) advocated for demonstration of proficiencies in evaluation domains through the compilation of a professional portfolio. Using Danielson’s FFT, Nielsen offered guidance for music teachers on how to demonstrate proficiency in each domain. Portfolio artifacts were used to guide the evaluator in identifying the teacher attributes of each domain. Nielsen’s premise was that it was the teacher’s responsibility to *display* proficiency within the context of the classroom rather than rely on the evaluator to *discover* proficiency. This concept was formulated to reduce the negative effect of an evaluator without content-area knowledge or experience. Nielsen wrote, “...the teacher portfolio provides a necessary link for the teacher and the evaluator to engage in meaningful conversations about effective teaching and to develop a mutual respect for the profession” (p. 69).

Previously mentioned factors affecting teacher evaluation included teacher practices, student learning measures, and the evaluation system used. Research has also described the role and impact the evaluator played in teacher evaluation.

The personality and/or disposition of the evaluator was a prominent theme in Todd's (2022) dissertation on the Emotional Intelligence (EI) level of a principal and its effect on job satisfaction and commitment. In this case study design mixed methods study, the principal took an Emotional Intelligence assessment and the participating teachers ($n = 27$) were asked to rate the principal on characteristics of EI that positively impacted job satisfaction. The results included relationships, empathy, optimism, and adaptability. Participants also listed traits that were not important to them including, emotional expression, self-esteem, sociability, and assertiveness. The results of the principal's responses and teachers' responses about the principal were then compared. Results of the comparisons between the principal's perceptions and teachers' perceptions varied with the principal's perception being higher or slightly lower depending on the category.

Trust in the personnel and process of evaluation was a recurring theme for Harris (2018) in his comparative case study. "The qualitative approach was appropriate for this study because it allowed insight to be gained about the thoughts and feelings of teachers and school leaders" (p. 32). His purpose was to compare the perceptions between administrators and music teachers of what was needed for professional growth. Participants included two administrator/evaluators and two music teachers. Harris was present during the observations and conducted interviews of both the evaluators and teachers revealing the following themes:

- Music teachers perceived professional development opportunities with content specific experts better cultivates their growth as compared to evaluation.
- Adequate resources play a vital role in cultivating the growth of music teachers.
- Trust in the execution of the evaluation process and evaluator cultivates teacher growth.

- Teacher age/experience plays a role in how they respond to feedback from leaders.

When well received, feedback cultivates teacher growth. (Harris, 2018, pp. 67-71)

Katz-Cote (2016) used narrative inquiry in her qualitative study to follow the journey of a teacher from new teacher to music supervisor. She transcribed conversations between herself and two other music supervisors on topics related to teaching and evaluating music, student growth, the transition from teacher to supervisor, and the need for community among supervisors. She referenced Papay (2012) and the need to have evaluators with content knowledge perform evaluations. She advocated for professional development for music teachers and music supervisors alike in teacher evaluations.

In an elaborately designed two-part study, Hirokawa (2013) studied differences in the evaluation of music teachers between administrators with different levels of music experience and music supervisors. Part I was a causal-comparative design where sixty-three administrators and music supervisors with differing backgrounds in music evaluated a teacher via a fourteen-minute video. Not surprisingly, the results showed a measurable difference in evaluation where the evaluators with the most musical knowledge (music supervisor) rated the teacher highest when compared to the administrator with some or no musical knowledge.

Part II included five participants in a quasi-experimental study that included a pre-test and post-test evaluation with training in between. The training consisted of a researcher-prepared 75-minute training video on music teacher evaluation. The post-test results showed an improvement in evaluation ratings that approached the scores of those with experience in music.

Kennedy (2021) examined the perspectives of administrators responsible for the evaluation of music educators in Texas. Semi-structured interviews were conducted in this qualitative study. Kennedy found that evaluators had confidence using the Texas Teacher

Evaluation and Support System (T-TESS) in general teacher evaluation and showed less confidence when evaluating music teachers. Research questions focused on how administrators were trained in the evaluation of music educators and their perspectives on using T-TESS for music teacher evaluation. This qualitative phenomenological study featured the interview of fifteen administrators. Kennedy identified characteristics such as good communication and classroom management were given more weight in place of specific content knowledge. His recommendations included professional development for administrators to include music teacher evaluation, the addition of personnel with music knowledge to assist with evaluations, and increased communication between the administrator and music teacher.

Formative and Summative Teacher Evaluation

Teacher evaluation has served multiple purposes and its usefulness has typically fallen into one of two categories: *formative* or *summative* evaluation (Malunda et al., 2016; Papay, 2012; Winters, 2021). Formative evaluation looks forward and can be used to inform professional development specific to the need revealed through the evaluation. It is prescriptive and has the goal of improving teacher effectiveness and practices. Formative evaluation, by nature, is a process that happens over a longer period of time. Even though research has shown a direct and positive link between formative evaluation and teacher effectiveness (Malunda et al., 2016), the amount of time and effort leads some already-too-busy administrators to over-rely on summative evaluations. Summative evaluation looks backward and can be used to impose a reward or penalty.

Malunda et al. (2016) studied teacher evaluation in Uganda in response to “the persistent criticisms about the deteriorating quality of teaching and learning in secondary schools in the country” (p.118). Other background concerns were a lack of proper lesson planning and use of outdated teacher-centered pedagogies rather than student-centered pedagogies. Their descriptive

cross-sectional survey included over 1,000 participants from ninety-five high schools and found that teacher evaluation was rarely formative and very inconsistent. They used the quality of pedagogical practices as the dependent variable and the independent variables were formative and summative evaluations. The survey consisted of three parts, including (a) demographics, (b) attitudes on teacher evaluation, and (c) pedagogical practices. Malunda et al. also interviewed evaluators and reviewed observation results for comparison with survey results.

A majority of the participants were between twenty and forty years old ($n = 664$, 71.1%) and male ($n = 644$, 69%). The results indicate that 83% of teachers surveyed possessed the requisite qualifications and 81% had been teaching for three or more years. Participants were asked to respond to items regarding teacher evaluations by indicating “agree,” non-committal,” or “disagree.” Slightly over half (54.5%) agreed that their evaluation was a fair assessment of their performance and 59% indicated their evaluation had a “great” impact on the way they teach. Malunda et al. (2016) concluded that “that formative teacher evaluation plays a significant role in enhancing the quality of pedagogical practices” and “summative teacher evaluation significantly contributed to increased quality of pedagogical practices” (p. 130).

Overland (2014) described these two uses of evaluation in terms of what the student knows or student outcomes (summative) and what the teacher does or teacher practices (formative). He noted the shift from earlier forms of evaluation where “...teachers might remember evaluations having been simple planned or surprise ‘walk-through’ observations by a principal or other arts faculty, followed by private discussions about ways to improve” (p. 56) to a teacher evaluation system more heavily influenced by value-added measures of student growth.

Ford and Hewitt (2020) called for the dismantling of the false dichotomy that requires the choice between either formative or summative evaluation. This false dichotomy led to a tension

between the evaluation purposes and an overreliance on one or the other. Instead, they advocated for the necessarily complementarian co-existence of the two to create an *integrated framework* utilizing the strengths of each evaluation mode. “To develop teacher evaluation systems and processes which are more balanced in terms of their consideration of both individual and collective needs, we need a framework which attempts to reconcile these two approaches” (Ford & Hewitt, 2020, p. 4). They sought to align formative evaluation with Self-determination theory (SDT) under the umbrella of individual/teacher interests. Summative evaluation was aligned with Stronge’s Improvement-oriented model under the umbrella of organizational/school interests. Their research found overlapping areas of mutual interest and articulated linkages between the two with implications for teacher evaluation. By assessing publicly available resources, their research investigated teacher evaluation in Hawaii and Washington, D.C. as two opposite ends of the formative/summative spectrum. Materials posted on official websites indicated that Hawaii favored a formative approach with the goal of improving teacher effectiveness. The Washington, D.C. resources indicated a more summative process with threats for underperforming teachers and lucrative rewards of up to \$25,000/year for high-performing teachers (Ford & Hewitt, 2020). From their research they recommended an integrated model with five distinct components: “compatible goals (individual and organization); two-way communication; supportive climate; technical rationality; and use of multiple data sources” (p. 9).

In a seminal article published in the *Harvard Review* (Spring 2012), Papay studied two means of measuring teacher performance – value-added models and standards-based observations – and advocated for their inclusion in both the summative determination of teacher effectiveness and the formative role in teacher development. The problem addressed in his article

echoed many already presented in Chapter I of this study and other sources cited here in Chapter II:

Over the past decade, consensus has been growing among teachers, administrators, and policy makers: teacher evaluation in the United States is broken and needs fixing. In school districts across the country, few teachers are evaluated regularly, and the evaluations that do occur are cursory. Not surprisingly, nearly all teachers succeed on these evaluations, and very few teachers are identified as unsatisfactory. These limitations have led to calls for reform, and districts across the country have struggled to identify and implement better evaluation systems. (Papay, 2012, p. 123)

His solution for mitigating the inherent problems with these measures was to identify and minimize bias. Standard-based observations are conducted through human observation which allows for subjective judgement. Value-added models remove subjectivity but may not account for contextual, cultural, or content-area differences. He also worked to increase reliability concerns that stemmed from use of multiple evaluators. Reliability has been shown to decrease with the use of multiple raters (inter-rater reliability), year-to-year differences in test results, and test-to-test differences in the same year. Validity can be affected by teacher practices and the inclusion of student growth measures outside the subject area. Papay (2012) concluded his study by reiterating the need for a teacher evaluation system that not only measured teacher effectiveness, but one that also supported teacher development. “Evaluations must provide teachers with a clear understanding not only of their current success or failure, but also of the practices they need to develop to become more successful with their students” (p. 138).

Accountability measured by summative evaluation provides the basis for merit increases for effective teachers, as well as disciplinary action for ineffective teachers. Decisions on merit

pay, promotion/advancement, retention, tenure, or other administrative decisions depend on appropriate and accurate teacher evaluation. Donaldson and Papay (2015) wrote that while these two functions of evaluation can complement each other, the potential for tension between them exists. Concerns raised in the study include the cost of teacher evaluation and the time and training of personnel to carry out the evaluations. The authors presented research indicating that some districts spent \$3,000-5,000 per teacher Peer Assistance and Review (PAR) evaluation system where teachers were trained to conduct evaluations in place of administrators. Cincinnati Public Schools reportedly spent between \$1.8 and 2.1 million per year on teacher evaluations.

Donaldson and Papay (2015) cited research supporting the premise that the increased focus on teacher evaluation has led to improved teacher practices and raised student scores while improving retention and helping to remove low performing teachers from the profession. One common criticism of teacher evaluation systems was the positive skewing of teacher effectiveness ratings. Donaldson and Papay cited Dee and Wyckoff (2013) who noted a teacher evaluation reform study in Washington, D. C. which showed a more typical curve with 16% of teachers receiving the highest rating and 15% receiving a low rating. Donaldson and Papay's conclusions were optimistic regarding the emphasis and impact of emerging teacher evaluation systems: "Emerging evidence suggests that new and enhanced evaluation systems may result in improved teacher performance, as measured by student achievement, and higher rates of selective attrition, thus addressing both of evaluation's purposes: development and accountability" (p. 188).

A review of the literature regarding teacher evaluation in both its formative and summative applications set the stage for current issues arising in teacher education: High-stakes teacher evaluation (HSTE) and evaluating teachers in non-tested grades and subjects (NTGS).

High-Stakes Teacher Evaluation

Robinson (2017) studied the impact of high-stakes teacher evaluation (HSTE) on retention and termination decisions. He described the two types of motivation that can be activated to bring about change in teacher practices: intrinsic and extrinsic.

Intrinsic motivators are things like achievement, recognition of competence, and increased levels of responsibility—characteristics that derive from the satisfaction of an individual doing the tasks associated with one's work—while extrinsic motivators take the form of compensation, promotions, merit pay, and other inducements—and are controlled primarily by others. (Robinson, 2017, p. 47)

The survey contained eleven open-ended items designed to collect qualitative data on teacher perceptions of HSTE. Participants were seven teachers from five states who had previously expressed interest in HSTE. Four were male and three female teachers with teaching experience ranging from five to thirty-three years in elementary ($n = 3$), middle school ($n = 3$), and high school ($n = 1$). The survey was distributed by email with follow-up emails and phone calls. Survey items included demographics, teaching context, and questions addressing “stress and pressure” and music-specific concerns associated with HSTE.

Robinson found teachers felt the pressure of HSTE impacted them causing them to experience perceptions of fake teaching, subject-matter inequity, hoop jumping, and a lack of support. Intrinsically motivated teachers who were evaluated under systems driven by extrinsic motivational beliefs experienced “major obstacles and challenges to their feelings of self-efficacy, professional and personal identity, and even continuation in the profession” (Robinson, 2017, p. 52).

Robinson noted that the participants who expressed these concerns had consistently been rated as “effective” or “highly effective” in their own evaluations. His recommendations included providing discipline-specific professional development opportunities for teachers subjected to HSTE, eliminating the use of VAMs, and returning the focus of teacher evaluation to formative evaluation with an emphasis on increasing teacher effectiveness through professional development.

High-stakes teacher evaluation (HSTE) can also have the unintended consequence of *performativity*. Shaw (2019) described a hypothetical situation where music teachers *teach-to-the-test* by creating measurable student learning objectives (SLOs) that will demonstrate student growth, but not necessarily enrich the students or create meaningful growth in the subject area. In high-stakes teacher evaluation, the goal is to receive a “good” evaluation. Pragmatism and performativity may seem the easiest solution to the problem of inaccurate measuring/evaluative tools and inexperienced evaluators. This unintended consequence produces teachers who are “highly qualified” and students who have “mastered” the learning outcomes but have likely not made real progress. Shaw (2019) advocated that music teachers “should consider asking for control of their documentation of student achievement and growth.” With this control comes the opportunity to create content-specialist-created and administrator-approved learning outcomes designed to accurately evaluate both student growth and teacher practices.

Lavigne (2014) addressed some of the intended and unintended outcomes of HSTE. With the inclusion of business models including results-driven ideologies into teacher evaluation, Race to the Top (RTTT, 2012) became a catalyst for increased pressure and more emphasis on HSTE. Lavigne analyzed RTTT award-winning states regarding their use of HSTE and “firing policies” and identified a notable difference between business use of high-stakes evaluation and that used

in education – a firing quota. Business models set a level (10%, 5%, 2%, etc.) for dismissals. The data on dismissal levels for ineffective teachers in schools using HSTE was less evident. Lavigne cited a rate of 2-4.5% in one of the two districts studied (D.C. public schools) and no mandate for dismissal in the other (Tennessee). Lavigne cited numerous other studies regarding the use of HSTE for dismissal including Berliner, 2014; Collins and Amrein-Beardsley, 2014; Lavigne and Good, 2013; and Roch et al., 2007.

Lavigne (2014) further discussed the identification of effective and ineffective teachers. When ineffective teachers were given opportunities to improve or be fired, questions of reliability of the evaluation criteria, percentage of ineffective teachers needing to be replaced, and availability of effective replacements remained unaddressed. Among the unintended outcomes of HSTE were teacher attrition, stress, and job satisfaction (also see Robinson, 2017). For many teachers, “high-stakes” equates to “high stress” and leads to some effective teachers leaving the profession. Lavigne’s conclusion was that “it remains unclear whether or not high-stakes teacher evaluation will meet the intended outcomes of a more effective teacher workforce and improved student achievement” (Lavigne, 2014, p. 22).

Not all responses to HSTE have been negative. The District of Columbia Public Schools implemented IMPACT, a teacher evaluation system that linked teacher performance with retention and pay. Arnot, Dee, and Wyckoff’s (2017) study of IMPACT pointed to two positive signs under this system: increased student learning and higher rate of under-performing teachers leaving. This controversial and heavy-handed policy had an almost immediate effect:

DCPS dismissed the majority of very low performing teachers and replaced them with teachers whose students did better, especially in math. Other low-performing teachers

were 50 percent more likely to leave their jobs voluntarily, and those who opted to stay improved significantly, on average, the following year. (p. 60)

The impact was also felt on the other end of the spectrum as high-performing teachers showed improvements, possibly as a result of efforts to earn financial incentives offered under the evaluation system.

The focus of DeSoto's (2018) quasi-experimental quantitative study on Hawaii's teacher evaluation was the relationship between HSTE and job satisfaction in the era of Race to the Top incentives. The study involved data from 248 schools and compared job satisfaction rates two years before the implementation of HSTE, at implementation, and two years after. The results were mixed, showing an increase in job satisfaction at the period of implementation of HSTE when compared to the prior measure. A surprise to the researcher was the drop in satisfaction following the implementation of HSTE even though it concurred with a pay raise. DeSoto concluded that the variation in levels of satisfaction was not directly related to the HTSE measures but rather possibly linked to some other intrinsic or systemic factors.

Bernard (2015) used a phenomenological interview approach to investigate the lived experience of music teacher evaluation in an environment of high-stakes teacher evaluation. Eight teachers and seven administrators were interviewed. Interview questions covered professional and personal background, experiences with evaluation, and interactions between administrator and teacher in the context of evaluation. The "results" of this phenomenological study were unique in that there was not an attempt to answer a question or solve a problem but merely to draw attention to the lived experiences of participants who were music teachers and administrators with a music teacher evaluation connection.

The necessity of teacher evaluation has been well established in other research, and the formative goals of improving teaching and summative goals of rating teachers' effectiveness in

relation to student learning are commonly accepted. As shown in this section, concerns arose about how these goals are accomplished and to what end they might also be employed. For instance, using summative evaluation to guide administrative decisions such as retention or dismissal, promotion or tenure decisions, or class assignments raise the level of impact on teachers.

Non-tested Grades and Subjects

Concern regarding the use of student growth in determining teacher effectiveness has been expressed due to the myriad of factors, both inside and outside school. These concerns were relevant even when the tested subjects and the subject taught aligned. An even more troubling concern was the use of student growth as measured by standardized tests and applied to teachers of non-tested grades and subjects (NTGS) – grade levels and subject areas not assessed by standardized tests.

The difficulty evaluating teachers of NTGS extended beyond the music classroom. While the focus of this writing was music teacher evaluation, music educators were not alone in their desire for accurate evaluation. Non-tested subjects included physical education, foreign languages, and the arts. To comply with early mandates, some districts divided teachers from the non-tested subjects into groups and added test results from available standardized tests in the evaluation of these teachers. The result was a teacher evaluation system based partially on student growth outside the content area of these teachers' subject area and control.

Parkes et al. (2015) studied the difficulty of evaluating teachers in NTGS due to the lack of accurate measures of student growth relative to the NTGS teachers. “The challenge of measuring student growth in areas such as music without access to standardized testing has prompted concern in the arts about the validity of assessment protocols” (Parkes et al., 2015, p.

24). The researchers created measurement instruments based on the Tennessee model but aligned with Virginia Standards of Learning. These instruments were used to gather data to be included in the portfolios which were reviewed by three master teachers. Artifacts included two samples from students representing beginning and end of the observation period. The study included evaluator training. The participating teachers reported positive response to this evaluation process, including a high level of professional and personal satisfaction with the validity of the student growth measures and their participation in the project.

Though ESSA expanded what constitutes a *core* subject to include “the arts, humanities, sciences and social sciences, in addition to English language arts and mathematics” (Jones & Workman, 2016, p.2), change takes time and progress seems slow. The evaluation of teachers in these newly included core subjects was not addressed in ESSA. Subjects are still defined in terms of “core” and “elective,” creating a hierarchy of importance, whether intended or not. Music teachers may feel isolated from these core subject teachers but face challenges similar to their colleagues in other NTGS.

Phillips et al. (2017) studied evaluation of high school physical education teachers in three school districts in a longitudinal study spanning twenty-one months and identified two relevant concerns. First, they raised concern regarding which student growth data would be used to evaluate teacher effectiveness; would student growth be related to physical education teaching practices? Second, they questioned whether classroom observation tools (Danielson’s *Framework for Teaching*) were designed to accommodate the uniqueness of a physical education teacher in their *classroom*. The participants were seven teachers (male = 6, female = 1) with between ten and thirty-four years of experience. The researchers interviewed administrators and teachers and conducted focus group discussions with students ($n = 36$). They also reviewed the

results of the teacher observations. The coded results revealed four main themes (a) change in curriculum and assessment, (b) the effect on administrators, (c) stakeholder apathy, and (d) increased department collaboration (Phillips et al., 2017). The student growth measures used in their evaluation were created by the physical education departments in order to maintain the validity of the results. The conclusion of the researchers was that changes occurred, but further study was needed to determine if the changes were meaningful or long-lasting.

As previously stated in this paper, measuring the effect of teacher practices on student learning in tested grades and subjects is commonly assessed by the results of standardized tests. The relationship between student success and teacher efficacy is examined using value-added modeling and student growth percentiles (SPG) (Gagnon et al., 2016). The linear nature of this relationship begins to deteriorate as the distance increases between the subject tested and the subject taught. The perceived appropriateness and accuracy of evaluation of NTGS teachers by these means is called into question.

For teachers not in these areas – those in the so-called ‘non-tested subjects and grades’ (NTSG) – the lack of vertical standardised achievement results largely precludes the use of VAM and SGP, and other techniques which purport to estimate teacher effectiveness using student outcomes had not been developed until much more recently. (Gagnon et al., 2016, p. 490)

Emert et al. (2013) outlined Pennsylvania’s efforts to address what was considered a “major challenge” in evaluating teachers of non-tested grades and subjects. The Educator Effectiveness model evaluation system was passed into law in 2012 and was implemented in 2013. Using this model, NTGS teachers were evaluated based on (a) building-level data (15%), (b) elective data (35%), and (c) direct observation (50%). The elective data included student

learning outcomes which allowed music teachers a means of demonstrating student achievement relative to course content.

In October 2019, the Pennsylvania State Education Association issued a news release regarding their teacher evaluation system. The statement summarized legislation to reform teacher evaluation that placed greater emphasis on teacher observations and reduced the weight of standardized tests and other measures outside the scope of the classroom teacher. Direct observations previously comprised 50% of the evaluation and under the new legislation, observations comprise 70% of the evaluation (PMEA, 2019).

With the foundation built upon a review of relevant literature in the areas of teacher evaluation, formative and summative evaluation, high-stakes teacher evaluation, and non-tested grades and subjects, literature regarding *music* teacher evaluation was reviewed.

Music Teacher Evaluation

Recognizing the difficulty in accurately evaluating teachers of subjects without standardized tests, Indiana music educators created their own rubric. The Music Teacher Effectiveness Rubric was developed by a task force formed by the Indiana Music Education Association (Gerrity, 2013). Existing evaluation rubrics were general and based on the underlying presumption that “good teaching is good teaching” regardless of the subject. The new rubric was designed to provide rubric-based examples of what good teaching or professional competencies look like in the music education context. It evaluated seven music teacher specific competencies:

- Teaches Comprehensively
- Engages Students in a Variety of Music Experiences
- Differentiates Instruction

- Provides for the Application of Musical Skills and Knowledge
- Utilizes Musically Appropriate Assessments
- Demonstrates a Commitment to Cross-Curricular Instruction
- Provides a Model for Professionalism

Overland (2014) offered a national clarion call to those concerned with accurate and effective music teacher evaluation. He reported that thirty-seven states revised their teacher evaluation systems toward the use of student growth as measured on achievement tests. A recurring theme in Overland's study was the difficulty of applying evaluation processes to non-tested or artistic classes resulting in teachers being evaluated on criteria beyond their pedagogical skills and student achievement in their classes.

Overland articulated a shift in paradigm from a more formative evaluation process focused on professional development, to the high-stakes and incentivized summative evaluation impacting bonus pay, retention, and tenure considerations for music teachers. He identified shortcomings in using VAMs to measure student growth as well as the difficulty using professionally developed teacher evaluation rubrics such as the *Danielson Framework for Teaching* and the *Marzano Teacher Evaluation Model*, both of which lacked the flexibility to accommodate music teacher evaluation (Overland, 2014). His article, published in the *Music Educators Journal* (2014), referenced teacher evaluation concerns and practices in Florida, Georgia, Illinois, Tennessee, Texas, South Carolina, New York, Delaware, and Ohio.

Concern for accurate music teacher evaluation is not a new phenomenon. In a seminal report, Taebel (1990) cited concerns over the "applicability of generic competencies to music teaching" and questioned the "qualifications of an observer or evaluator who is not trained in music" (p. 50). His major points of concern from thirty years ago remain the subject of research

and most pressing questions for current music teachers: “by what standards and by whom will we be evaluated?” Taebel cited studies from Georgia, Florida, Texas, and Alabama where music teachers consistently scored lower than average on competencies used to evaluate all teachers. His concern was that the impact of inaccurate evaluation of music teachers could lead to a move away from best practices toward demonstration of generic competencies, loss of confidence in music teaching ability, and cynicism toward the evaluation process. He cited the Music Educators National Conference (MENC, 1988) “Policy Statement on Music Teacher Evaluation” which recommended the modification of evaluation programs to adequately evaluate music teachers, the design of special assessment instruments to identify the special competencies required by music educators, and evaluation personnel who are knowledgeable in music.

Taebel (1990) concluded that the explanation for consistently low-scoring performance by music teachers could be attributed to one of two reasons: (a) music teachers lack the professional competencies of their non-music teaching colleagues, or (b) the specific teaching behaviors required for success in music education are not recognized by the evaluation system.

Brophy (1993) continued the conversation by reiterating the same concern regarding teacher evaluation that does not acknowledge the unique classroom environment and professional competencies necessary for success in music education: “When state legislatures are determining generic criteria for teacher competencies, there seems to exist an underlying assumption that all subjects are taught in the same manner” (Brophy, 1993, p. 1). The purposes of his study were to examine relevant research and identify issues relating to music teacher evaluation. Using data from the nine studies he reviewed, he articulated eight “domains” considered relevant for the effective evaluation of music educators:

- personal characteristics, as evidenced through student-teacher, teacher-faculty, and teacher-administration interactions
- musical competence and performance skill, as evidenced when teaching a lesson
- effective use of nonverbal strategies, such as modeling/demonstrating
- effective use of verbal strategies, such as eliciting performance from students
- classroom management, including discipline, group participation, and the creation of a positive learning environment
- effective planning for concept learning and aesthetic appreciation across a wide age span
- an objective assessment of teaching style based on empirically supported criteria
- appropriate professional development activities should be recommended. (Brophy, 1993)

Brophy cited MENC efforts toward more accurate music teacher evaluation by including a list of characteristics relevant to teacher effectiveness in the music classroom and instructions for raters who evaluate music teachers (Appendix A).

Robinson (2015) blasted the teacher evaluation world by stating, “That we have a national education policy called Race to the Top should tell us everything we need to know about how wrong this approach truly is. Education is not a race” (Robinson, 2015, p. 20). Instead, he offered that education is a process that when functioning properly, everyone wins. In “The Inchworm and the Nightingale: On the (Mis)use of Data in Music Teacher Evaluation,” Robinson (2015) reviewed national policies on teacher evaluation, reviewed research on teacher evaluation, and made suggestions regarding music teacher evaluation.

Robinson referred to the improvement (formative) and accountability (summative) purposes of teacher evaluation and the “uncoupling” of their dual purposes in favor of teacher evaluation weighted more heavily on accountability which carried a punitive nature rather than an instructive nature. He presented three extreme examples of the impact of this shift:

- In 2010, every teacher at Central Falls High School in Rhode Island was fired because of low student test scores.
- In February 2014, students at a school in Tennessee were given after-school detention and extra homework assignments after scoring poorly on practice tests for a state math exam.
- In December 2011, New York City school officials announced that twelve schools had been placed on a school closing list, primarily because of low scores on state tests.

Robinson (2015) cited the National Association for Music Education (NAfME, 2011) position statement on teacher evaluation. Below are excerpts from his summary of the statement relating to measures of student achievement and music teacher evaluation (for complete statement, see Appendix A):

- Measures of student achievement used in teacher evaluation must be based on student achievement that is directly attributable to the individual teacher, in the subject area taught by that teacher
- Successful music teacher evaluation must include measures of music student achievement in the areas of creating, performing, and responding and must avoid using school-wide measures other than those directly associated with music achievement (NAfME, 2011)

Salvador and Krum (2019) shared the concern that best practices in some music classes “are not necessarily the same as the teacher behaviors described by indicators on teacher evaluation rubrics” (p. 146). The problem created by a lack of discipline-specific evaluation rubrics is compounded if the personnel lacks content knowledge or experience in evaluating classes outside of core subjects. They found that “standardized teacher evaluation can influence instruction in ways that do not align with best practices in music education” (p. 150). The study offered a series of fictitious vignettes depicting the evaluation of music teachers in their context. The vignettes depicted many “best practices” that fell outside the parameters of routine teacher evaluation. The response from the teachers was to move away from music best practices, refrain from taking chances with new pedagogies in favor of simpler and more traditional teaching in line with what the evaluator “expects” to see in the classroom. Salvador and Krum voiced concern about the impact of improper music teacher evaluation on the inclusivity and cultural responsiveness needs of the students in the classroom. Music teachers are tempted to remain in the *status quo* instead of taking the risks “necessary for the messy, creative, innovative work of equity, inclusion, and cultural responsiveness” (p. 140). Salvador and Krum continue,

It is also unrealistic to think that an evaluator who is not familiar with the philosophy and practices of a content area (let alone a specific musical praxis) would be able to interpret teaching and learning interactions without being welcomed into those philosophies and practices through discussion, observation, and participation. (p. 150)

The attention on music teacher evaluation produced local, state-level, and national results. In Minnesota, educational entities including the state teacher’s union, Minnesota Department of Education, Minnesota Music Educators Association, along with local teachers unions and other educational organizations worked on how new education law was understood

and applied. While not always directly working together, the synergy of working toward the same goal resulted in the creation of The Collaboration, Growth, and Evaluation Model which included teacher practices (45%), student engagement (20%) as measured by student survey and reflection, and student learning and achievement (35%) as measured by standardized assessments and/or teacher designed student learning goals (Orzolek, 2014).

Minnesota Music Educators Association (MMEA) produced articles, research, and professional development opportunities for teachers to focus on the creation of these student learning goals. These professional development opportunities were designed to create an environment of collaboration and support that promoted learning and growth.

Advocacy for music as a valued part of education in America has centered around the National Association for Music Education (NAfME). Their efforts to draw attention to issues affecting music education, in general, and music teacher evaluation, specifically have been referenced in some of the literature reviewed thus far. In 2016, NAfME updated two resources to help teachers and administrators with music teacher evaluation. *Workbooks for Building and Evaluating Effective Music Education* were created for both general music and for ensembles. The workbooks were designed to support a teacher's evaluation of their program or support the teacher and administrator in teacher evaluation. They included guidance on measuring student achievement including evaluation worksheets with teacher evaluation rubrics compatible with evaluation systems designed by Danielson, Marzano, and Marshall. These resources put the content and context of the music classroom into language used by administrators for teacher evaluation. Specific music examples helped the administrator identify what they were looking for in classroom observations (NAfME, 2016).

Use of Value-added Measures

Value-added measures (VAMs) employ a sophisticated statistical model designed to translate student test scores into teacher effectiveness estimates (Haertel, 2013). The tone for discussion can be set by the previously cited NEA white paper which noted, "...numerous studies have concluded that value-added methodology is neither fair enough, nor reliable enough, nor valid enough to be used as a basis for high-stakes decisions about teachers" (NEA, 2010, p.8). Arguably one of the most problematic and most researched aspects of teacher evaluation is the use of Value-added Measures (VAMs) in determining teacher effectiveness. As the role and influence of the federal government has grown, the impact has been felt in teacher evaluation. Donaldson and Papay (2015) reported forty-one states requiring the value-added measure of student test scores as part of the teacher evaluation equation. Problems with using VAMs for teacher evaluation have been well documented (Amrein-Beardsley & Geiger, 2019; Donaldson & Papay, 2015; NEPC, 2019) and are even more troubling when used to evaluate teachers of NTGS (Croft & Buddin, 2015; Gilbert, 2016; Hall & Marion, 2016; Phillips et al., 2017; Zirkel et al., 2015).

In the opening chapter of Hewitt and Amrein-Beardsley's (2016) *Student Growth Measures in Policy and Practice Intended and Unintended Consequences of High-stakes Teacher Evaluations*, the authors traced the history of VAMs entrance into the teacher evaluation process beginning with their use as a metric for school effectiveness to their gradual use as a metric for teacher effectiveness. Hewitt and Amrein-Beardsley noted that by 1998, VAMs were used to account for up to 8% of a teacher's effectiveness in Tennessee. The weight VAMs increased to 35-50% during Race to the Top (2009) and resulted in the award of \$500 million to Tennessee. The authors cited studies linking teacher effectiveness to student learning and the

lasting effect of this relationship in positive (for effective teachers) and negative (for ineffective teachers) ways. The intended result of VAMs used in teacher evaluation was the removal of ineffective and advancement of effective teachers.

To better calculate teacher effectiveness, Tennessee developed an evaluation system that incorporated portfolios demonstrating student growth. Their goal was to address a nation-wide problem of how to include student growth in subjects where standardized tests are not offered. The artifacts evaluated in the portfolios were directly connected to the content of the fine arts classes (and later, foreign language and physical education classes). Robelen (2013) described portfolio content including videos of students sight-reading or reciting poetry, photos of student work and samples of research and other assignments.

The Tennessee Fine Arts Growth Measures System (TVAAS) used portfolios to account for 35% of the evaluation with observations comprising 50% and 15% consisting of standardized tests or other assessments (Robelen, 2013). The piloted multi-year, multi-discipline project used portfolios for teacher evaluation in fine arts, first grade, physical education, pre-K/kindergarten, and world languages. The system provided a means for teachers to document their practices and the impact they are having on student learning. Using TVASS, teachers were given an individual growth score that correlated to their contributions to student learning and reflected a more accurate and personal evaluation experience. The peer-reviewed portfolio scores were consistent with those obtained by observation and teachers experienced a slight increase in overall evaluation score. Teachers also reported increased satisfaction with the use of portfolios in their evaluation.

There has long been a premise that evaluating teacher performance and holding teachers accountable for student test results would “motivate teachers to work harder and smarter and

help attract and retain only those who are successful” (Harris & Herrington, 2015, p. 72). Harris and Herrington (2015) raised pertinent questions regarding VAMs in teacher evaluation. Did teachers believe VAMs improved their performance or led to more shallow teaching-to-the-test? Did sanctioning or firing low-performing teachers encourage or discourage creativity and motivation in high-performing teachers? Did the use of VAMs erode trust between teachers and administrators? Did the focus on individual performance minimize efforts to collaborate with other teachers and build the team for the benefit of the school as a whole?

Although Harris and Herrington (2015) raised more questions than answers, they envisioned a crossroad ahead where schools, districts, or states would decide to (a) reduce the intensity (accountability) of current systems, (b) maintain intensity but reduce or eliminate the use of VAMs, or (c) modify the current system while maintaining the intensity and the use of VAMs, observations, and other measures.

Croft and Buddin (2015) and Zirkel et al. (2015) seemingly argued against each other while making the same point: VAMs were an inappropriate means of teacher evaluation for NTGS. In each study, the authors cited relevant court cases where the legality of the use of VAMs was called into question. In the cited cases, the court found the use of VAMs was legal but unfair (Zirkel et al., 2015). They refer to Haertel’s (2013) parameters for limited use of VAMs for NTGS teacher evaluation if certain conditions could be met:

- Scores based on sound, appropriate student tests
- Comparisons limited to homogeneous teacher groups
- No fixed weight- flexibility to interpret VAM scores in context for each individual case
- Users well trained to interpret scores

- Clear and accurate information about uncertainty (e.g., margin of error) (Haertel, 2013)

Shuler (2012), a NAFME past president, agreed in opposing the use of VAMs in music teacher evaluation, stating, “Evaluating teachers of music based on test scores in subjects they do not teach is not only unfair; it is absurd” (p. 9). He reiterated the importance of solving this educational dilemma and encouraged teachers to work with local music education associations and NAFME to create solutions, rather than complain or do nothing.

Finding sensible solutions to these questions is important to all teachers, but particularly important to ensure fair evaluation of educators who teach subjects—including music—that are not measured by standardized state tests (Shuler, 2012, p. 9).

In a scathing review of VAMs used for teacher evaluation, Close and Amrein-Beardsley (2018) wrote, “When rethinking their teacher evaluation systems under ESSA, education leaders will do well to look to the past for guidance on what not to do — and what to do better” (p. 15). They cited four “main blunders” of the system: (1) inconsistent findings, (2) overlooking evidence of bias, (3) allowing people to game the system, and (4) avoiding transparency. They suggested a system comprised of multiple measures, formative evaluation, and teacher ownership and engagement in the process. Not content to stop with a critique of a problem, the authors offered suggestions for improving teacher evaluation including, (a) using multiple measures to evaluate teachers, (b) designing teacher evaluation systems that emphasize formative uses, and (c) engaging teachers throughout the process of creating and refining these systems (Close & Amrein-Beardsley, 2018).

Summary

The purpose of this study was to investigate high school music teachers' levels of satisfaction with the evaluative practices used by administrators and music supervisors in the Commonwealth of Virginia. Virginia high school music teachers were surveyed to collect data to determine their levels of satisfaction with their evaluation (RQ 1) and what, if any, factors related to their levels of satisfaction (RQ 2). The satisfaction level was measured using a researcher-constructed *Teacher Evaluation Satisfaction Survey (TESS)* completed by high school music teachers in Virginia. The data collected were used to answer the research questions presented in this study.

This review of the literature examined teacher evaluation, the purposes of teacher evaluation, high-stakes teacher evaluation, non-tested grades and subjects, music teacher evaluation, and the use of value-added measures. While researching these topics, the researcher cited several related state-specific studies which were included to provide a deeper understanding of impact teacher evaluation is having across the country. Many of the studies reviewed included the use of surveys to collect data. Surveys allow characteristics, attitudes, or behaviors to be identified in a small group – the sample – to provide insight into those same characteristics in the larger group – the population where identifying the same features would be excessively time-consuming, expensive, or impossible to collect.

Over the past two decades, Federal education law, specifically No Child Left Behind (NCLB, 2002), Race to the Top (RttT, 2009), and Every Student Succeeds Act (ESSA, 2015) has drawn attention to school performance, student outcomes, and teacher effectiveness. Teacher effectiveness, as measured by teacher evaluation systems, has long been an important part of the education process due to the relationship between teacher effectiveness and student growth

(Blumberg, 1985; Marzano, 2001 & 2011). The literature reviewed in this chapter showed a shift in the focus of teacher evaluation from its formative function of improving teacher effectiveness through professional development to a more summative function of improving teacher effectiveness by ranking, rating, removing ineffective teachers, and rewarding effective teachers. With its guidelines, mandates, and incentives, education law and policy has impacted teachers in both intended and unintended ways. Research has revealed the use of high-stakes teacher evaluation and use of value-added measures have created a high-stress environment concerning teacher evaluation (Lavigne, 2014; Robinson, 2015 & 2017; Shaw, 2019). There has been some evidence of success in the intended outcome of removing ineffective teachers from the profession, but the impact on effective teachers includes increased stress and attrition (Adnot et al., 2017).³

The most common concern in the reviewed literature was the use of value-added measures in determining teacher effectiveness. The appropriateness of VAMs in determining teacher effectiveness has been called into question in research presented here and beyond. Relevant to the current study was the use of VAMs in measuring teacher effectiveness for non-tested grades and subjects (NTGS). Many value-added measures have used standardized test scores to measure student growth. The student growth measures were combined with other factors to determine teacher effectiveness. This led to teachers in NTGS being evaluated by criteria outside their control (Gagnon et al., 2016; Phillips et al., 2017).

National attention on teacher evaluation resulted in responses from organizations including, the National Education Association (NEA), the National Association for Music

³ September 29, 2022, the National Education Policy Center posted a newsletter entitled “Is teacher stress a policy problem” referencing federal education law, job satisfaction, high-stakes teacher evaluation, and teacher stress. <http://nepc.colorado.edu/publication/newsletter-stress>

Education (NAfME), and the National Education Policy Center (NEPC) in the form of white papers, position statements, and newsletters (Appendix A). State level responses include research directed at state-specific concerns (DeSoto, 2018; Emert et al., 2013; Gerrity, 2013; Kennedy, 2021; Orzolek, 2014; Overland, 2014; Parkes et al., 2015; Robelen, 2013). State music education associations produce resources in collaboration with NAfME via the *Music Educators Journal*. Access has been available to NAfME members or through library resources such as Sage Journals (<https://journals.sagepub.com>).

Additionally, this review of the literature presented dissertations from universities across the country and from various disciplines. Degrees sought included Doctor of Philosophy (Ph.D.), Doctor of Education (Ed.D.), and Doctor of Music Education (D.M.E.). The variety in discipline, geography, and type of degree brought attention to the broad impact teacher evaluation had on the teaching profession. The dissertations reviewed in this chapter highlighted the interest in and concern regarding teacher evaluation, high-stakes teacher evaluation, music teacher evaluation, and job satisfaction (Bernard, 2015; DeSoto, 2018; Harris, 2018; Hirokawa, 2013; Katz-Cote, 2016; Kennedy, 2021; Todd, 2022).

CHAPTER III: METHOD

Restatement of the Purpose

The purpose of this study was to investigate high school teachers' levels of satisfaction with the evaluative practices used by administrators and music supervisors in the Commonwealth of Virginia. Virginia high school music teachers ($n = 76$) were surveyed to collect data to determine their levels of satisfaction with their evaluation (RQ 1) and what, if any, factors were related to their levels of satisfaction (RQ 2). The satisfaction level was measured using a researcher-constructed *Teacher Evaluation Satisfaction Survey (TESS)*. The data collected were used to answer the following research questions:

1. What are the levels of satisfaction, among high school music teachers, using contemporary evaluative practices?
2. Are there specific factors that relate to the levels of satisfaction with teacher evaluation?

Participants

In order to accurately assess the levels of satisfaction with teacher evaluation, high school music teachers ($n = 76$) were surveyed, using the *Teacher Evaluation Satisfaction Survey (TESS)*. The participants were limited to teachers who met the following parameters: (a) high school teachers, (b) music teachers, and (c) in the Commonwealth of Virginia. The assumption was that music teachers, as a whole, shared the same obstacles in teacher evaluation without regard to music area, such as, band, choir, orchestra, general music, etc. There was no limitation to participate based on school size, location, or school type (public or private).

Participant Selection

Participants were identified through their membership in the Virginia Music Educators Association (VMEA). The stated purpose of VMEA is “to provide leadership and professional development to ensure quality music education” (www.vmea.com). VMEA has over 3,200 members. Participants were invited to participate in the study through the VMEA Newsletter and mailing lists. The invitation to participate included a link to the online survey. Response rate determination was hampered due to the nature of the invitations. There was no process available to determine the number of members who saw the invitation and chose not to participate.

In accordance with IRB protocols, participant privacy was ensured using an anonymous online survey where no identifying information or personal data was collected by the researcher.

Participant Demographics

Study participants were not selected based on demographic considerations apart from those previously mentioned (high school music teachers in Virginia), however, demographic information was collected to determine if there was any relationship between or among the demographic identifiers and the levels of satisfaction with teacher evaluation. The demographic items used for analysis were age, sex, Race/Ethnicity, highest level of education, years of service, school type, school size, school setting, teacher evaluation performed by, and subject taught. A descriptive analysis of participant demographics was presented in Chapter IV: Results.

The Survey Instrument

Among the various types of studies available for use in a quantitative study are descriptive surveys, experiments, causal-comparative studies, and correlational studies (Lodico et al., 2010). Surveys are questionnaires designed to gather “opinions, beliefs, or perceptions about a current issue from a large group of people” (p. 207). Gall et al. described the use of

questionnaires and interviews in educational research “to collect data about phenomena that are not directly observable: inner experience, opinions, values, interests, and the like (Gall et al., 2007, p. 228).

Creswell (2014) defined the purpose of surveys, “to generalize from a sample to a population so that inferences can be made about some characteristic, attitude, or behavior of this population” (p. 157). Creswell recommended the inclusion of an introductory cover letter, the items, and closing instructions. Among the items are demographic information as well as attitudinal, behavioral, and factual items (statements/questions). Establishing validity and reliability were of primary concern in the design of the survey. A panel of experts was used to confirm validity and pilot testing confirmed reliability (Froehlich & Frierson-Campbell, 2013). These and other researchers have outlined procedures for creating, administering, and collecting surveys.

The survey was constructed using Qualtrics Survey Software, a web-based product designed to create surveys and produce reports. The survey instrument for the current study was constructed to collect data on music teacher perceptions on the levels of satisfaction with the methods used in their evaluation. The purpose of survey research is to gather data from a sample group that represents the larger population, in this case, high school music teachers in the Commonwealth of Virginia. If validity and reliability were high, the findings could be appropriately generalized to the whole (Gall et al., 2007). The participants responded to prompts using a Likert-type scale that consisted of “strongly disagree,” “somewhat disagree,” “neither disagree nor agree,” “somewhat agree,” or “strongly agree.”

The *Teacher Evaluation Satisfaction Survey (TESS)* was constructed by the researcher in three parts (Appendix B): Part 1- Demographics (Table 1); Part 2- Process, Personnel, and

Product of Teacher Evaluation (Table 2); and Part 3- Open-ended Items (Table 3). Part 1 gathered demographic data including age, gender, Race/Ethnicity, highest level of education, and years of service. Each section in Part 2 contained prompts answered on a 5-point Likert-type scale addressing the participants experience and understanding of the process, personnel, and product of their evaluation. Part 3 provided the opportunity for participants to respond, in their own words, to prompts regarding characteristics of positive or negative evaluations, “good” or “bad” evaluators, and advice for new teachers, evaluators, and the researcher.

Table 1. *TESS Part 1*

Participant Demographics

1. Age (select one):

20-30

31-40

41-50

51-60

61-70

71 or older

Prefer not to say

2. Sex:

Male

Female

Prefer not to say

Table 1. Cont.

Participant Demographics
3. Race/Ethnicity:
White/Caucasian
Asian - Eastern
Asian - Indian
Hispanic
African-American
Native-American
Mixed race
Other _____
Prefer not to say
4. Highest level of education received:
Bachelor's
Master's
Doctorate
Other _____
5. Subject taught (select all that apply):
Band
Choir
Orchestra
General Music
Theory
Piano
Guitar
Other _____

Table 1. Cont.

Participant Demographics	
6. Years of service:	
	0-5
	6-15
	15-25
	25 or more
7. School type:	
	Public
	Private
	Charter
8. School size:	
	1,000 or less
	1,001-2,000
	2001 or more
9. School setting:	
	Urban
	Suburban
	Rural
10. Teacher Evaluations are performed by:	
	Building Administrator
	Fine Arts Supervisor
	Other_____

Part 2 used *Likert* scale responses for participants to respond to their perceptions on the *process, personnel, and product* of their evaluation (Table 2):

Table 2. TESS Part 2

Teacher Evaluation Process
1. I understand the evaluation process for my position.
2. I know the criteria upon which I am being evaluated.
3. I am given time to prepare for my evaluation.
4. My evaluation highlights teacher practices.
5. My evaluation includes observation(s).
6. My evaluation includes student growth measures in my content area.
7. My evaluation includes student growth measures outside my content area.
8. The criteria used are appropriate for HS music teacher evaluation.
9. I am satisfied with the process of evaluation at my school.

Teacher Evaluation Personnel
1. I know the person who will be evaluating me.
2. I trust the person evaluating me to be objective.
3. My evaluator is competent to evaluate my teaching in my context.
4. My evaluator has experience in music supervision.
5. My evaluator is experienced with non-tested grades and subjects.
6. I am satisfied with the personnel who conduct my evaluations.

Teacher Evaluation Product
1. My evaluation provides an accurate summary of my performance.
2. My evaluation impacts my job security.
3. My evaluation impacts advancement decisions.
4. My evaluation impacts pay raise decisions.
5. My evaluation has led to improvements in my teaching.
6. My evaluation has led to improvements in student learning.
7. I am satisfied with the product of my evaluation.

Part 3 contained open-ended or free-response questions allowing participants to share their lived-experiences in their own words (Table 3):

Table 3. *TESS* Part 3

Open-ended Items
1. What advice would you give to new music teachers to your campus regarding music teacher evaluations?
2. What advice would you give to evaluators to your campus regarding music teacher evaluation?
3. Is there anything you would like to share with me, as a researcher, that was not covered in this survey on music teacher evaluation?

Prior to survey distribution, approval from the Institutional Review Board (IRB) was sought. Participation in this study posed minimal risk to the participants and the IRB exempted it from further review.

Survey Pilot

During the development of the survey, several steps were taken to confirm both the reliability and validity of the instrument. A preliminary draft of the survey was reviewed by music education experts at two local universities including the dissertation advisor. Minor changes were made based on the recommendations and comments of this focus group. Changes included providing adequate space for open questions, changes in wording, and a review of content to support the validity of the survey.

Following the initial presentation to the expert panel, and using the revisions suggested, the survey instrument was piloted by a convenience sample of Virginia music educators with similar profiles to the desired group of survey participants. The survey link was emailed to the pilot group and the link anonymized the responses. The pilot also provided an opportunity to test the distribution method and link to the survey. The pilot participants provided feedback to the

researcher including amount of time spent on the survey, ease of use, and clarity of the survey. Once the recommendations from the focus group of music education higher education professionals and pilot survey participants were evaluated, the survey was modified for distribution.

Survey Distribution

The Virginia Music Educators Association (VMEA) is a state-affiliate of the National Association for Music Education (NAfME) serving music educators in Virginia. Advocacy, professional development, and news are among the services they provide their members. An email inquiry requesting assistance with the survey distribution was sent to the VMEA Office. The VMEA Newsletter was chosen as the vehicle for the invitation to participate. The link took potential participants to Qualtrics for a description, instructions, and the survey.

Data Analyses

For each of the “parts” of the survey, the data were analyzed to accurately understand the purpose of the data in serving to answer the research questions.

Part 1

A descriptive analysis of the demographic data revealed the frequencies, means, standard deviations, and percentages for each category. The results were reported in tables in Chapter IV: Results. The ability to generalize the data from the sample in this study to the population requires that the participants’ demographic characteristics are comparable to the demographic characteristics of the larger population.

Part 2

Part 2 responses were collected by using a 5-point Likert scale. The instructions are presented in Table 4.

Table 4. Likert-type question responses

Instructions for Part 2

For the following statements, please select the response most accurately reflecting your agreement- (1) strongly disagree, (2) somewhat disagree, (3) neutral, (4) somewhat agree, (5) strongly agree.

Before analysis, the responses were organized using a Microsoft Excel spreadsheet. This allowed for easier management of the data and transfer into IBM's SPSS, version 27 for statistical analysis. The results from Part 2 were analyzed by two-way analysis of variance (ANOVA) to determine if significant differences existed in levels of satisfaction with teacher evaluation between groups (age, sex, years of experience, etc.). Multiple linear regression analysis was used to identify independent variables (components of process, personnel, and product of teacher evaluation) and their predictive relationship to the dependent variable of *satisfaction*.

Part 3

The open-ended items in the survey were analyzed to identify common themes. The process of coding was used to identify recurring themes present in the responses. Creswell (2014) articulated three steps for coding: (1) organize the data, (2) read all the data, and (3) code the data (p. 197). The responses to the open-ended items were collected and read multiple times by the researcher to identify common themes. As themes were identified, they were color-coded into groups. Each group was reported by its theme in Chapter IV: Results. Coded themes were evaluated to determine what role they played in answering the research questions and how they impacted levels of satisfaction with teacher evaluation. Results were reported in table format in Chapter IV: Results.

Summary

The purpose of this study was to investigate high school teachers' levels of satisfaction with the evaluative practices used by administrators and music supervisors in the Commonwealth of Virginia. Virginia high school music teachers were surveyed to collect data to determine their levels of satisfaction with their evaluation (RQ 1) and what, if any, factors related to their levels of satisfaction (RQ 2). The satisfaction level was measured using a researcher-constructed *Teacher Evaluation Satisfaction Survey (TESS)* completed by high school music teachers in Virginia. The data collected were used to answer the research questions presented in this study.

This chapter discussed the methods used during the study. The process used for participant selection was discussed including the participant attributes required for inclusion of the study. Since the context of the study is high school music teacher evaluation in Virginia, it was determined that participants needed to (a) be a high school teacher, (b) be teaching music, and (c) be teaching in Virginia. Participant demographics were collected in Part 1 of the *Teacher Evaluation Satisfaction Survey (TESS)*.

Survey construction, distribution, and data analysis was discussed. The researcher used a focus group including the dissertation advisor and other higher education music education professionals. The survey was piloted using a convenience sample of music educators with characteristics similar to those of the participants sought for the survey. Changes based on the information from the focus group and pilot survey responses were considered for incorporation to the survey. The Virginia Music Educators Association was contacted and facilitated the survey distribution through their newsletter. Data collected by the survey were analyzed using descriptive analysis (Part 1), analysis of variance (ANOVA) and multiple linear regression (Part 2), and coding for recurring themes (Part 3).

CHAPTER IV: RESULTS

The purpose of this study was to investigate high school teachers' levels of satisfaction with the evaluative practices used by administrators and music supervisors in the Commonwealth of Virginia. Virginia high school music teachers were surveyed to collect data to determine their levels of satisfaction with their evaluation (RQ 1) and what, if any, factors related to their levels of satisfaction (RQ 2). The satisfaction level was measured using a researcher-constructed *Teacher Evaluation Satisfaction Survey (TESS)*, completed by high school music teachers in Virginia. The data collected were used to answer the following research questions:

1. What are the levels of satisfaction, among high school music teachers, using contemporary evaluative practices?
2. Are there specific factors that relate to the levels of satisfaction with teacher evaluation?

The survey was constructed in three parts: Part 1- Demographics; Part 2- Process, Personnel, and Product of Teacher Evaluation; and Part 3- Open-ended Items. Part 1 gathered demographic data including age, gender, Race/Ethnicity, highest level of education, and years of service. Each section in Part 2 contained prompts answered on a 5-point Likert-type scale addressing the participants experience and understanding of the process, personnel, and product of their evaluation. Part 3 provided the opportunity for participants to respond, in their own words, to prompts regarding characteristics of positive or negative evaluations, "good" or "bad" evaluators, and advice for new teachers, evaluators, and the researcher.

Descriptive Analyses of Data

Demographics

Age

The age range of participants included all ages of professional activity following completion of teaching licensure (20-30, 31-40, 41-50, 51-60, 61-). The most frequently observed age range was 31-40 ($n = 26$, 34.2%) followed by 20-30 ($n = 19$, 25.0%) and 41-50 ($n = 18$, 23.7%). Two participants were 71 or older (2.6%) and nine were 51-60 (11.8%).

Sex

The sex of the participants was almost evenly divided (Female: $n = 37$, 49.3% and Male: $n = 36$, 48.0%) with two preferring not to say. For comparison, the National Center for Education Statistics (NECS) reported 64% female and 36% male teachers for secondary schools.

Race/Ethnicity

The most frequently observed category of Race/Ethnicity was White/Caucasian ($n = 61$, 80.3%). Other participants were African-American ($n = 4$, 5.3%), Hispanic ($n = 8$, 10.5%), and Native American and Asian – Eastern ($n = 1$, 1.3%), with one participant preferring not to say.

Education Level

The most frequently observed response to highest level of education was Master's ($n = 42$, 55.3%) followed by Bachelor's ($n = 22$, 29.0%) and Doctorate ($n = 8$, 10.5%). The remaining responses ($n = 4$, 5.3%) included Education Specialist, Administration and Supervision certificate, and "progress in completing a Master's degree."

Years of Service

The most frequently observed response for years of service was 6-15 ($n = 29$, 38.2%) followed closely by 15-25 ($n = 23$, 30.3%) and 0-5 ($n = 13$, 17.1%). The least frequent range was

25 or more years of experience ($n = 11, 14.5\%$). Frequencies and percentages are presented in Table 5.

Table 5. Demographic Information

Variable	<i>n</i>	%
Age		
31-40	26	34.2
20-30	19	25.0
41-50	18	23.7
51-60	9	11.8
71 or older	2	2.6
Prefer not to say	1	1.3
Gender		
Female	37	49.3
Male	36	48.0
Prefer not to say	2	2.7
Race/Ethnicity		
White/Caucasian	61	80.3
Hispanic	8	10.5
African-American	4	5.3
Asian - Eastern	1	1.3
Native-American	1	1.3
Highest level of education received		
Master's	42	55.3
Bachelor's	22	29.0
Doctorate	8	10.5
Other	4	5.3

Table 5. Cont.

Variable	<i>n</i>	%
Years of Service		
6-15	29	38.2
15-25	23	30.1
0-5	13	17.1
25 or more	11	14.5

The teaching context was reported using responses regarding school type, size, setting, and subject(s) taught. The most frequently observed school type was Public ($n = 66$, 88.0%). The most frequently observed category of school size was 1,001-2,000 ($n = 33$, 43.4%). The most frequently observed category of school setting was Suburban ($n = 47$, 61.8%) followed by Rural ($n = 20$, 26.3%) and Urban ($n = 9$, 11.8%). Frequencies and percentages are presented in Table 6.

Table 6. Teaching Context

Variable	<i>n</i>	%
School type		
Private	8	10.7
Public	66	88.0
School size		
1,001-2,000	33	43.4
2,001-3,000	14	18.4
1,000 or less	23	30.3
3,001 or more	6	7.9
School setting		
Suburban	47	61.8
Urban	9	11.8
Rural	20	26.3

Participants reported “subject taught (select all that apply)” which yielded a total number of responses ($n = 144$) greater than the number of participants due to teaching contexts in which more than one subject is taught. The most frequently observed categories of “subject taught” were Band ($n = 33, 22.9\%$) and Choir ($n = 3, 22.8\%$). In addition to the choices listed on the survey, participants also added other subjects taught, including, music appreciation, theatre, music technology, percussion, and handbells. Frequencies and percentages are presented in Table 7.

Table 7. Teaching Context – subject(s) taught

Variable	<i>n</i>	%
Subject taught		
Choir	30	20.8
Band	33	22.9
General	12	8.3
Music	13	9.0
Orchestra	16	11.1
Theory	12	8.3
Piano	12	8.3
Guitar	7	4.9
Other	9	6.3

As referenced in Chapters I and II, the individual conducting the evaluation can impact evaluation outcomes for music teachers. Evaluation personnel and number of evaluations per year were reported. Overwhelmingly, Building Administrator was the most frequently observed category ($n = 69, 92.0\%$). The frequency of evaluation was relatively consistent with most participants being evaluated 1-3 times per year ($n = 40, 81.6\%$). Frequencies and percentages are presented in Table 8.

Table 8. Teaching context – personnel and frequency

Variable	<i>n</i>	%
Evaluation personnel		
Building Administrator	33	94.3
Other	2	5.7
Frequency of evaluation		
1-3 per year	40	81.6
Less than 1 per year	3	6.1
More than 3 per year	6	12.2

Teacher Evaluation Process

The first section of Part 2 of the survey investigated the *process* of the evaluation including open-ended items to gather information not included in the Likert scale responses. The 5-point Likert-type scale allowed for responses from 1 = strongly disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree, and 5 = strongly agree.

When responding to their understanding of the evaluation process, participants overwhelmingly ($n = 54$) either somewhat agreed ($n = 25, 44.6\%$) or strongly agreed ($n = 29, 51.8\%$). This positive response rate ($M = 4.5$) was among the highest average scores in the section. Similarly positive responses ($M = 4.3$) were reported regarding knowledge of criteria for evaluation with only three responding, “strongly disagree” and one responding “somewhat disagree.” When responding to the prompt, “my evaluation highlights teacher practices,” a sizeable majority of participants either somewhat agreed ($n = 24, 42.9\%$) or strongly agreed ($n = 20, 35.7\%$). “Strongly disagree” was the most frequent response ($n = 34, 64.2\%$) when responding to “my evaluation includes student growth measures outside my content area.” Conversely, the most frequent responses to “my evaluation includes student growth measures in my content area” were “somewhat agree” ($n = 18, 32.7\%$) and “strongly agree” ($n = 20, 36.4\%$).

Observations were included in a majority of the respondents' evaluations ($n = 37, 88.1\%$) and almost half ($n = 26, 46.4\%$) "strongly agreed" that they were given time to prepare for their evaluations.

The average (M) scores for appropriateness of criteria for evaluation ($M = 3.4$) and overall satisfaction with the process of evaluation ($M = 3.5$) are the lowest scores in this section. Overall satisfaction responses were "strongly disagree" ($n = 5, 8.9\%$), "somewhat disagree" ($n = 4, 7.1\%$), "neither agree nor disagree" ($n = 18, 32.1\%$), "somewhat agree" ($n = 14, 25.0\%$), and "strongly agree" ($n = 15, 26.8\%$). Additional variables were considered later in this chapter to identify any correlations impacting these responses. Frequencies and percentages are presented in Table 9.

Table 9. Teacher Evaluation Process

Variable	<i>M</i>	<i>SD</i>	<i>n</i>	Strongly Disagree (%)	Somewhat disagree (%)	Neither agree nor disagree (%)	Somewhat Agree (%)	Strongly agree
I understand the process	4.5	0.6	56	$n = 0$ (0%)	$n = 1$ (1.8%)	$n = 1$ (1.8%)	$n = 25$ (44.6%)	$n = 29$ (51.8%)
I know the criteria	4.4	1.0	55	$n = 3$ (5.5%)	$n = 1$ (1.8%)	$n = 0$ (0%)	$n = 19$ (34.6%)	$n = 32$ (58.2%)
My evaluation highlights teacher practices	4.0	1.1	56	$n = 3$ (5.4%)	$n = 2$ (3.6%)	$n = 7$ (12.5%)	$n = 24$ (42.9%)	$n = 20$ (35.7%)
My evaluation includes student growth measures outside my content area	1.9	1.3	53	$n = 34$ (64.2%)	$n = 5$ (9.4%)	$n = 6$ (11.3%)	$n = 4$ (7.6%)	$n = 4$ (7.6%)
My evaluation includes observations	4.5	0.9	42	$n = 1$ (2.4%)	$n = 0$ (0%)	$n = 4$ (9.5%)	$n = 8$ (19.1%)	$n = 29$ (69.1%)
The criteria used are appropriate	3.4	1.3	56	$n = 4$ (7.1%)	$n = 13$ (23.2%)	$n = 11$ (19.6%)	$n = 14$ (25.0%)	$n = 14$ (25.0%)

I am given time to prepare	4.0	1.3	56	<i>n</i> = 4 (7.1%)	<i>n</i> = 5 (8.9%)	<i>n</i> = 6 (10.7%)	<i>n</i> = 15 (26.8%)	<i>n</i> = 26 (46.4%)
My evaluation includes student growth measures in my content area	3.7	1.4	55	<i>n</i> = 8 (14.6%)	<i>n</i> = 3 (5.5%)	<i>n</i> = 6 (10.9%)	<i>n</i> = 18 (32.7%)	<i>n</i> = 20 (36.4%)
I am satisfied with the process of my evaluation	3.5	1.2	56	<i>n</i> = 5 (8.9%)	<i>n</i> = 4 (7.1%)	<i>n</i> = 18 (32.1%)	<i>n</i> = 14 (25.0%)	<i>n</i> = 15 (26.8%)

Participants responded to the prompt “my evaluation includes student growth measures in my content area” with a variety of teacher provided growth measures:

- Measures “chose by me each year”
- SMART goal data
- Teacher provided results of teacher created material
- Data based on pre-test, mid-year, post-test- music literacy, vocabulary, and score reading
- Sight reading
- Assessment of student growth
- Singing of an Art Song
- My choosing – I pick Sight Reading
- Largely meaningless ‘data’ contrived to please the above musically illiterate Arts Administrator

When providing additional responses to the prompt “my evaluation includes student growth measures in my content area,” participants offered far fewer but more general responses such as “basic knowledge” which appear uninformative to this study.

Following the Likert scale responses about the process of evaluation, participants were prompted to “list some factors, practices, or characteristic that have a positive impact on your

levels of satisfaction with music teacher evaluation” and “list some factors, practices, or characteristic that have a negative impact on your levels of satisfaction with music teacher evaluation.” These open-ended questions gave the participants the opportunity to elaborate on topics from the Likert scale responses in a personal and specific way relevant to their context and experiences. The responses were coded by the researcher and grouped or clustered around common themes appearing in their responses. A summary of these coded responses is presented in Table 10. The complete listing of participants’ responses is included in Appendix C.

Table 10. Factors, practices, or characteristic that have a positive impact on your levels of satisfaction with music teacher evaluation

<i>Positive impact</i>	<i>Frequency</i>
Teacher input in evaluation	12
Respect, support, and relationship with evaluator	12
Level of understanding of music	7

The most frequent response was related to the ability of the music teacher to apply their content-specific filter to the evaluation process. The opportunity to set content-specific measurable goals and have input on selecting the measures for evaluation rose to the top of the list of factors having a positive impact on the levels of satisfaction with evaluation. The coding of participant responses also revealed a strong relational component to the evaluation process. Participants “appreciated” and “really appreciated” the “efforts,” “support,” “respect,” and being valued by their evaluators. An evaluator’s knowledge/experience or lack thereof also factored into the positive impact on satisfaction with evaluation. A summary of these coded responses is presented in Table 11. The complete listing of participants’ responses is included in Appendix C.

Table 11. Factors, practices, or characteristic that have a negative impact on your levels of satisfaction with music teacher evaluation

<i>Negative impact</i>	<i>Frequency</i>
Evaluator’s lack of knowledge/experience in music	17
Evaluations that do not “fit” music	15
Lack of feedback/input from music teacher	8

Not surprisingly, some of the themes revealed by the coding of the responses to characteristics having a positive impact on satisfaction were indirectly and conversely present in characteristics having a negative impact on satisfaction with evaluation. The most frequently mentioned theme was the evaluator’s lack of knowledge of music and music evaluation. Teacher input including student growth measures, performance objectives, and other shared goals and data were cited as negatively impacting satisfaction with evaluation if not included. Participants also referenced the lack of music-specific evaluations by referring to “one size fits all” or evaluations “not specific to music teaching.”

Teacher Evaluation Personnel

The second section of Part 2 asked participants to respond to prompts regarding the personnel conducting their evaluations. The most frequently observed category of “I know the person who will be evaluating me” was “strongly agree” ($n = 47, 83.9\%$). A majority of participants responded, “strongly agree” to their trust in the objectivity of their evaluator ($n = 28, 50.9\%$), however five responded either “strongly disagree” ($n = 3, 5.5\%$) or “somewhat disagree” ($n = 2, 3.6\%$). The low average score ($M = 3.0$) showed a lack of confidence in the competence of the evaluator with roughly half responding either “strongly disagree” ($n = 10, 17.9\%$) or “somewhat disagree” ($n = 15, 26.8\%$) and the other half responding, “strongly agree” ($n = 9, 16.1\%$) or “somewhat agree” ($n = 17, 30.1\%$). Even lower average scores were reported when

responding to “experience in music supervision” of the evaluator ($M = 1.9$) and similarly when responding to the evaluator’s experience with NTGS ($M = 2.8$). Over 75% of participants responded, “strongly disagree” ($n = 28, 50.0\%$) or “somewhat disagree” ($n = 14, 25.0\%$) when asked if their evaluator had experience in music supervision.

The final prompt in this section addressed satisfaction with evaluation personnel and had an average response of ($M = 3.4$). Just over half of the participants reported “somewhat agree” ($n = 13, 23.6\%$) or “strongly agree” ($n = 15, 27.3\%$) regarding their satisfaction with the personnel conducting their evaluations. The remaining responses were “neither disagree nor agree” ($n = 13, 23.6\%$), “somewhat disagree” ($n = 8, 14.6\%$), and “strongly disagree” ($n = 6, 10.9\%$).

Frequencies and percentages are presented in Table 12.

Table 12. Teacher Evaluation Personnel

Variable	<i>M</i>	<i>SD</i>	<i>n</i>	Strongly Disagree (%)	Somewhat disagree (%)	Neither agree nor disagree (%)	Somewhat Agree (%)	Strongly agree
I know the person who will be evaluating me	4.8	0.4	56	$n = 0$ (0%)	$n = 0$ (0%)	$n = 1$ (1.8%)	$n = 8$ (14.3%)	$n = 47$ (83.9%)
I trust the person to be objective	4.2	1.1	55	$n = 3$ (5.5%)	$n = 2$ (3.6%)	$n = 5$ (9.1%)	$n = 17$ (30.9%)	$n = 28$ (50.9%)
My evaluator is competent to evaluate my teaching in my context	3.0	1.4	56	$n = 10$ (17.9%)	$n = 15$ (26.8%)	$n = 5$ (8.9%)	$n = 17$ (30.4%)	$n = 9$ (16.1%)
My evaluator has experience in music supervision	1.9	1.2	56	$n = 28$ (50.0%)	$n = 15$ (25.0%)	$n = 6$ (10.7%)	$n = 6$ (10.7%)	$n = 2$ (3.6%)
My evaluator is experienced with NTGS	2.8	1.4	56	$n = 14$ (25.0%)	$n = 10$ (17.9%)	$n = 14$ (25.0%)	$n = 10$ (17.9%)	$n = 8$ (14.3%)
I am satisfied with the personnel who conduct my evaluations	3.4	1.3	55	$n = 6$ (10.9%)	$n = 8$ (14.6%)	$n = 13$ (23.6%)	$n = 13$ (23.6%)	$n = 15$ (27.3%)

Following the Likert scale responses for the personnel of evaluation, participants were prompted to share “some characteristics found in a ‘good’ evaluator.” Their responses were coded to identify common themes or characteristics of a “good” evaluator. A summary of the most frequent themes is presented in Table 13. The complete listing of participants’ responses is included in Appendix C.

Table 13. "Good" evaluator

<i>Characteristics found in a “good” evaluator</i>	<i>Frequency</i>
Understanding, objective, willing to learn	28
Content knowledge	14
Good communication skills	10

Participants were nearly unanimous in their agreement that “understanding” and a “willingness to learn” was key to being a “good” evaluator. The cluster of responses in this group was almost double the next most frequent response of “content knowledge” which has been referenced previously in this study. Understanding, flexibility, objectivity, patience, and a willingness to learn the unique environment of a music classroom were all mentioned in the responses. Content knowledge, while a critical and often discussed aspect of successful music teacher evaluation seems to be less influential to satisfaction in evaluation than a willingness to understand and learn about the music classroom. Good communication skills was the third most frequent group of responses reported by the participants. Communication in scheduling, conducting, and providing feedback in the evaluation process were all referenced.

Teacher Evaluation Product

As articulated in the introduction, the purposes of teacher evaluation include academic (student outcome) and professional (teacher practice) goals. The third section of Part 2 of the survey focused on the Teacher Evaluation Product. Study participants responded to prompts

designed to describe whether or not they believe their evaluations provide an accurate summary of their performance, impacts job security, impacts advancement and pay, improves their teaching practices, and improves student learning.

The most frequently observed category of “my evaluation provides an accurate summary of my performance” was “somewhat agree” ($n = 28, 50.0\%$) comprising half of the responses. Only four (7.1%) of the participants “strongly disagreed” with this prompt. When responding to “my evaluation impacts my job security,” over half of the participants responded either “somewhat agree” ($n = 22, 39.3\%$) and “strongly agree” ($n = 12, 21.4\%$). The effect of evaluations on advancement decisions showed similar results with “somewhat agree” ($n = 20, 35.7\%$) and “strongly agree” ($n = 9, 16.1\%$). The participants responded that pay raises were less affected by their evaluations with “strongly disagree” ($n = 26, 47.3\%$) as the most frequent response.

In Chapter I: Introduction, two important goals of evaluation were referenced: teacher practices and student learning. The prompts “my evaluation has led to improvements in my teaching” and “my evaluation has led to improvements in student learning” showed lower than expected average scores of $M = 2.9$ and $M = 2.7$. When asked if evaluations led to improvements in their teaching, participants “strongly disagreed” ($n = 11, 19.6\%$) and “somewhat disagreed” ($n = 12, 21.4\%$) more than they “strongly agreed” ($n = 9, 16.1\%$) and “somewhat agreed” ($n = 11, 19.6\%$). Regarding improvements in student learning related to evaluation, participants “strongly disagreed” ($n = 13, 23.6\%$) and “somewhat disagreed” ($n = 12, 21.8\%$) more than they “strongly agreed” ($n = 6, 10.9\%$) and “somewhat agreed” ($n = 9, 16.4\%$).

Not surprisingly, the average score ($M = 3.1$) showed lower satisfaction regarding the product of evaluation than the satisfaction with personnel ($M = 3.4$) and process ($M = 3.5$) of

evaluation. Only eight participants (14.3%) “strongly agreed” that they were satisfied with the product of their evaluation. “Somewhat agree” ($n = 15, 26.8\%$) and “neither agree nor disagree” ($n = 19, 33.9\%$) were the most frequent responses in this category. Frequencies and percentages are presented in Table 14.

Table 14. Summary Statistics Table for Interval and Ratio Variables

Variable	<i>M</i>	<i>SD</i>	<i>n</i>	Strongly Disagree (%)	Somewhat disagree (%)	Neither agree nor disagree (%)	Somewhat Agree (%)	Strongly agree
My evaluation provides an accurate summary of my performance	3.5	1.1	56	$n = 4$ (7.1%)	$n = 8$ (14.3%)	$n = 9$ (16.1%)	$n = 28$ (50.0%)	$n = 7$ (12.5%)
My evaluation impacts my job security	3.6	1.1	56	$n = 3$ (5.4%)	$n = 5$ (8.9%)	$n = 14$ (25.0%)	$n = 22$ (39.1%)	$n = 12$ (21.4%)
My evaluation impacts advancement decisions	3.3	1.2	56	$n = 7$ (12.5%)	$n = 5$ (8.9%)	$n = 15$ (26.8%)	$n = 20$ (35.7%)	$n = 9$ (16.1%)
My evaluation impacts pay raise decisions	2.3	1.5	55	$n = 26$ (47.3%)	$n = 8$ (14.6%)	$n = 9$ (16.4%)	$n = 4$ (7.3%)	$n = 8$ (14.6%)
My evaluation has led to improvements in my teaching	2.9	1.4	56	$n = 11$ (19.6%)	$n = 12$ (21.4%)	$n = 13$ (23.2%)	$n = 11$ (19.6%)	$n = 9$ (16.1%)
My evaluation has led to improvements in student learning	2.7	1.3	55	$n = 13$ (23.6%)	$n = 12$ (21.8%)	$n = 15$ (27.3%)	$n = 9$ (16.4%)	$n = 6$ (10.9%)
I am satisfied with the product of my evaluation	3.1	1.3	56	$n = 9$ (16.1%)	$n = 5$ (8.9%)	$n = 19$ (33.9%)	$n = 15$ (26.8%)	$n = 8$ (14.3%)

Part 3, “In their own words,” was three open-ended questions created to allow the participants to speak to new music teachers regarding teacher evaluation, to speak to evaluators regarding the evaluation of music teachers, and to speak to the researcher. The responses listed

below were grouped by theme in the analysis section of the results. Their responses are presented in Tables 15, 16, and 17.

Participants were asked, “What advice would you give to new music teachers to your campus regarding music teacher evaluations?” Their responses were coded to identify common themes or characteristics in their advice. A summary of the most frequent themes is presented in Table 15. The complete listing of participants’ responses is included in Appendix C.

Table 15. Advice for new teachers

<i>What advice would you give to new music teachers to your campus regarding music teacher evaluations?</i>	<i>Frequency</i>
Relax, be yourself, just teach	26
Ask questions/communicate with evaluators about expectations	10
Connect with other music teachers for mentoring/coaching	9

The advice offered in the form of participant responses ranged from “take them with a grain of salt” to “they are meaningless” to the more specific “have an agenda on the board and make sure you state goals for each rehearsal chunk and then remember to circle back at the end to check in with student to see how they feel about the concepts taught.” Some of the more encouraging comments indicated that evaluations were “encouraging, positive, helps to improve my teaching,” and can “sometimes provide helpful insight for improvement.”

The participants were then asked, “What advice would you give to evaluators to your campus regarding music teacher evaluation?” Their responses were coded to identify common themes or characteristics in their advice. A summary of the most frequent themes is presented in Table 16. The complete listing of participants’ responses is included in Appendix C.

Table 16. Advice for evaluators

<i>What advice would you give to evaluators to your campus regarding music teacher evaluations?</i>	<i>Frequency</i>
Express empathy and a desire to understand the music classroom	17
Content knowledge	15
Specific feedback on for improvement	10

Responses were grouped by coding of themes into three main categories (Table 16). The most frequent response ($n = 17$) related to the person/personality or relational posture of the evaluator. The most common theme was “understand/understanding,” and responses encouraged a relational approach where evaluators sought to learn and understand teaching in the music content in contrast to “just checking boxes to appease higher admin requirements.” A second theme ($n = 15$) emerged in the coding process highlighting a desire for “content knowledge.” Specifically, responses listed evaluators with experience in music education and/or the arts as very important. The final theme revealed by the coding process ($n = 10$) related to the specifics of the evaluation process and good feedback, measures of growth, and good communication from the evaluators regarding the process and expectations.

The last question in the open-ended item section allowed the participants to share with the researcher on any topic not covered on the survey. The answers centered around those provided to previous questions. The main theme was the need for evaluators who are experienced in music or at a minimum, music teacher evaluation. Suggestions included training for administrators the use of fine arts supervisors for music teacher evaluation.

Research Questions

The descriptive analysis of the participant demographics in Part 1, the survey content questions in Part 2, and the open-ended items in Part 3 created the data from which the statistical analyses were reported to answer the research questions addressed in this study.

1. What are the levels of satisfaction, among high school music teachers, using contemporary evaluative practices?
2. Are there specific factors that relate to the levels of satisfaction with teacher evaluation?

Research Question 1

“What are the levels of satisfaction, among high school music teachers, using contemporary evaluative practices?” Levels of satisfaction of evaluation were measured in three prompts in Part 2 of the survey. Responses to participants’ levels of satisfaction with the process, personnel, and product of teacher evaluation were measured. A summary is shown in Table 17 followed by discussion.

Table 17. Reported levels of satisfaction with evaluation

Variable	<i>M</i>	<i>SD</i>	<i>N</i>	Strongly Disagree (%)	Some-what disagree (%)	Neither agree nor disagree (%)	Some-what Agree (%)	Strongly agree
I am satisfied with the process of my evaluation	3.5	1.2	56	<i>n</i> = 5 (8.9%)	<i>n</i> = 4 (7.1%)	<i>n</i> = 18 (32.1%)	<i>n</i> = 14 (25.0%)	<i>n</i> = 15 (26.8%)
I am satisfied with the personnel who conduct my evaluations	3.4	1.3	55	<i>n</i> = 6 (10.9%)	<i>n</i> = 8 (14.6%)	<i>n</i> = 13 (23.6%)	<i>n</i> = 13 (23.6%)	<i>n</i> = 15 (27.3%)
I am satisfied with the product of my evaluation	3.1	1.3	56	<i>n</i> = 9 (16.1%)	<i>n</i> = 5 (8.9%)	<i>n</i> = 19 (33.9%)	<i>n</i> = 15 (26.8%)	<i>n</i> = 8 (14.3%)

Teacher Evaluation Process

The mean reported levels of satisfaction with “the process of my evaluation” was $M = 3.5$. Over half the participants somewhat agreed ($n = 14, 25.0\%$) or strongly agreed ($n = 15, 26.8\%$) responding to the prompt, “I am satisfied with the process of my evaluation.” The levels of satisfaction were compared between demographic characteristics.⁴ An analysis of variance (ANOVA) was run to determine whether there were significant differences in satisfaction with teacher evaluation process by category: age, sex, Race/ethnicity, highest level of education, years of service, school type, school size, school setting, or the person performing the evaluation. The ANOVA was examined based on an alpha value of .05. The results of the ANOVA were not significant, $F(21, 30) = 0.55, p = .918$, indicating the differences in satisfaction with the teacher evaluation process were similar across all demographic categories.

Teacher Evaluation Personnel

The mean reported levels of satisfaction with “the personnel who conducted my evaluation” was $M = 3.4$. Slightly over half the participants somewhat agreed ($n = 13, 23.6\%$) or strongly agreed ($n = 15, 27.3\%$) responding to the prompt, “I am satisfied with the personnel who conduct my evaluations.” The levels of satisfaction were compared between demographic characteristics.⁵ An analysis of variance (ANOVA) was run to determine whether there were significant differences in satisfaction with teacher evaluation personnel by category: age, sex, Race/ethnicity, highest level of education, years of service, school type, school size, school setting, or the person performing the evaluation. The ANOVA was examined based on an alpha

⁴ The demographic data presents the responses from 76 participants. All data for Part 2 and Part 3, and all resulting finds are calculated to reflect the 56 participants who responded fully to the survey.

value of .05. The results of the ANOVA were not significant, $F(20, 30) = 0.89, p = .598$, indicating the differences in satisfaction with the teacher evaluation personnel were similar across all demographic categories.

Teacher Evaluation Product

The mean reported levels of satisfaction with “the product of my evaluation” was $M = 3.1$. Less than half the participants somewhat agreed ($n = 15, 28.8\%$) or strongly agreed ($n = 8, 14.3\%$) responding to the prompt, “I am satisfied with the product of my evaluation.” The levels of satisfaction were compared between demographic characteristics. An analysis of variance (ANOVA) was run to determine whether there were significant differences in satisfaction with teacher evaluation product by category: age, sex, Race/ethnicity, highest level of education, years of service, school type, school size, school setting, or the person performing the evaluation. The ANOVA was examined based on an alpha value of .05. The results of the ANOVA were not significant, $F(21, 30) = 1.29, p = .257$, indicating the differences in satisfaction with the teacher evaluation product were similar across all demographic categories.

Averaging the three categories of satisfaction (means of the means) to determine a grand mean to serve as the overall levels of satisfaction with teacher evaluation (RQ1) produces a result of $M = 3.3$ where just over half of the participants report satisfaction with their teacher evaluation. A Levels of satisfaction Composite scale was calculated consisting of the three previously mentioned categories: teacher evaluation process, teacher evaluation personnel, and teacher evaluation product. A Cronbach alpha coefficient of .85 indicated good reliability and consistency between scores. Table 18 presents the results of the reliability analysis.

Table 18. Reliability Table for Satisfaction Composite

Scale	No. of Items	α	Lower Bound	Upper Bound
Levels of satisfaction Composite	3	.85	.79	.91

Note. The lower and upper bounds of Cronbach's α were calculated using a 95.00% confidence interval.

Research Question 2

“Are there specific factors that relate to the levels of satisfaction with teacher evaluation?” Before investigating the possibility of various factors’ effects on satisfaction with teacher evaluation, a Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in satisfaction with evaluation between the levels of various demographic data. The Kruskal-Wallis test is a non-parametric alternative to the one-way ANOVA and does not share the ANOVA's distributional assumptions (Conover & Iman, 1981).

Demographic Effect

While the demographic data presents the responses from 76 participants, all data for Part 2 and Part 3, and all resulting finds are calculated to reflect the 56 participants who responded fully to the survey. When investigating the potential effect of age, sex, and Race/ethnicity, the results of the Kruskal-Wallis test were not significant based on an alpha value of .05, indicating that the mean rank of satisfaction with the product of evaluation was similar for each level of these demographic characteristics. The results of the Kruskal-Wallis analyses are presented in Table 19.

Table 19. Results of Kruskal-Wallis analyses

Category	Analysis
Age	$\chi^2(5) = 3.81, p = .577$
Sex	$\chi^2(1) = 0.28, p = .597$
Race/ethnicity	$\chi^2(3) = 0.80, p = .849$

When investigating the potential effect of level of education and years of experience, the results of the Kruskal-Wallis test were not significant based on an alpha value of .05, indicating that the mean rank of satisfaction with the product of evaluation was similar for each level of these demographic characteristics. The results of the Kruskal-Wallis analyses are presented in Table 20.

Table 20. Results of Kruskal-Wallis analyses

Category	Analysis
Level of education	$\chi^2(2) = 1.85, p = .396$
Years of experience	$\chi^2(3) = 2.41, p = .492$

When investigating the potential effect of the teaching context including school type, school size, school setting, and evaluation personnel, the results of the Kruskal-Wallis test were not significant based on an alpha value of .05, indicating that the mean rank of satisfaction with the product of evaluation was similar for each level of these demographic characteristics. The results of the Kruskal-Wallis analyses are presented in Table 21.

Table 21. Results of Kruskal-Wallis analyses

Category	Analysis
School type	$\chi^2(1) = 1.22, p = .269$
School size	$\chi^2(3) = 3.35, p = .340$
School setting	$\chi^2(2) = 0.82, p = .664$
Evaluation Personnel	$\chi^2(1) = 0.03, p = .864$

After ruling out the effect of demographic characteristics on levels of satisfaction with the product of evaluation, multiple linear regressions were conducted to assess the effect of each independent variable (represented by questions) within each measure of satisfaction: process (Q13.1 – 13.8), personnel (Q15.1 – 15.5), and product (Q18.1 – 18.6). The results from the initial regression analysis (Appendix D) identified questions with responses not providing a significant

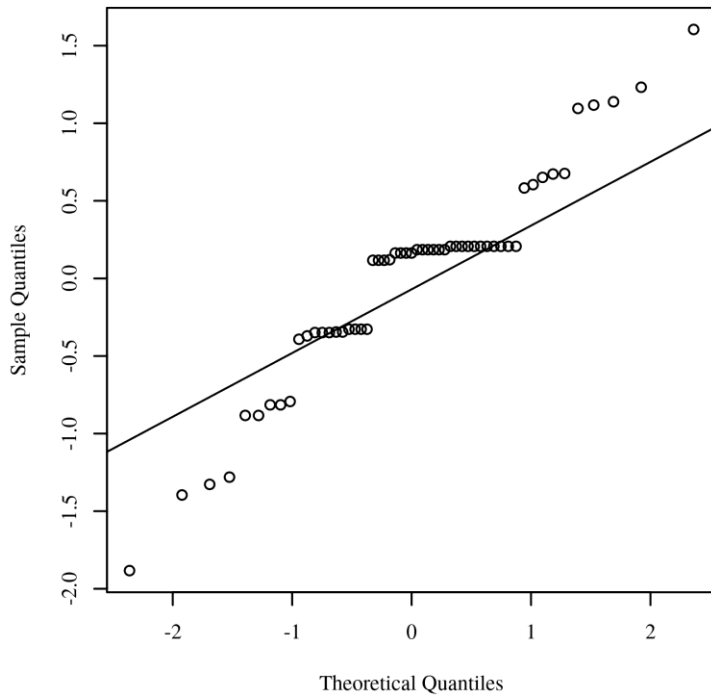
effect. Those questions were removed to minimize multicollinearity and another regression analysis was conducted to determine the predictive relationship between significant questions on the levels of satisfaction with teacher evaluation.

Teacher Evaluation Process

A linear regression analysis was conducted and the results were that “my evaluation highlights teacher practices” (Q13.4) and “the criteria used are appropriate for HS music teacher evaluation” (Q13.8) significantly predicted satisfaction with the process of evaluation (Q13.9).

The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot (DeCarlo, 1997). For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Figure 1 presents a Q-Q scatterplot of the model residuals.

Figure 1. Q-Q scatterplot for normality of the residuals for the regression model



Since the linear regression considered more than one independent variable, Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. High VIFs indicate increased effects of multicollinearity in the model. VIFs greater than 5 are cause for concern, whereas VIFs of 10 should be considered the maximum upper limit (Menard, 2009). All predictors in the regression model have VIFs less than 10. Table 22 presents the VIF for each predictor in the model.

Table 22. Variance Inflation Factors for Q13.4 and Q13.8

Variable	VIF
Q13.4	1.47
Q13.8	1.47

The results of the linear regression model were significant, $F(2,52) = 60.82, p < .001, R^2 = .70$, indicating that approximately 70.05% of the variance in satisfaction with the process of evaluation (Q13.9) was explainable by Q13.4 and Q13.8. Responses to “my evaluation highlights teacher practices” (Q13.4) significantly predicted satisfaction, $B = 0.51, t(52) = 4.92, p < .001$. Responses to “the criteria used are appropriate” (Q13.8) significantly predicted satisfaction, $B = 0.47, t(52) = 5.34, p < .001$. Table 23 summarizes the results of the regression model.

Table 23. Results for Linear Regression with Q13.4 and Q13.8 predicting Q13.9

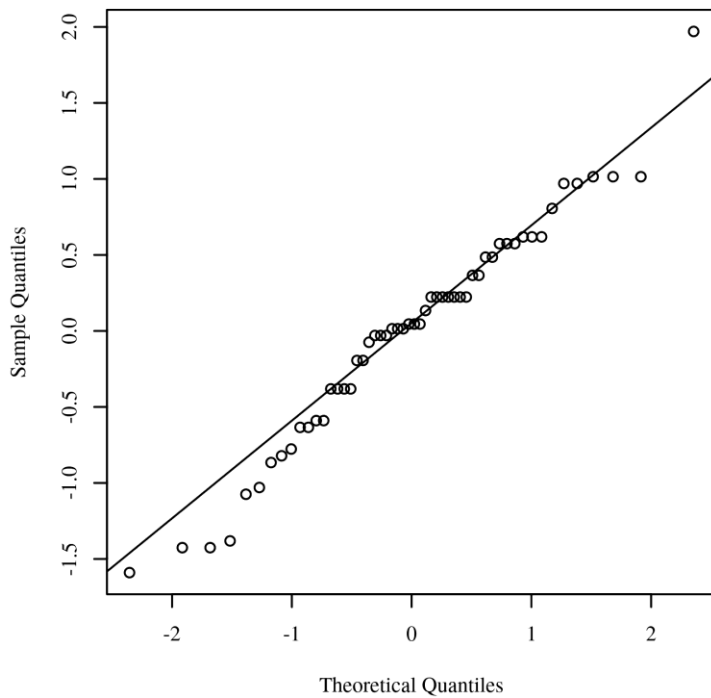
Variable	<i>B</i>	<i>SE</i>	95.00% CI	β	<i>t</i>	<i>p</i>
(Intercept)	-0.10	0.36	[-0.83, 0.63]	0.00	-0.27	.786
Q13.4	0.51	0.10	[0.30, 0.72]	0.45	4.92	< .001
Q13.8	0.47	0.09	[0.29, 0.64]	0.49	5.34	< .001

Teacher Evaluation Personnel

A linear regression analysis was conducted and the results were that “I trust the person evaluating me to be objective” (Q15.2) and “my evaluator is experienced with NTSG” (Q15.5) significantly predicted satisfaction with evaluation personnel (Q15.6).

The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution. Figure 2 presents a Q-Q scatterplot of the model residuals.

Figure 2. Q-Q scatterplot for normality of the residuals for the regression model



Since the linear regression considered more than one independent variable, Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. All predictors in the regression model have VIFs less than 10. Table 24 presents the VIF for each predictor in the model.

Table 24. Variance Inflation Factors for Q15.2 and Q15.5

Variable	VIF
Q15.2	1.51
Q15.5	1.51

The results of the linear regression model were significant, $F(2,51) = 53.17, p < .001, R^2 = .68$, indicating that approximately 67.59% of the variance in satisfaction with the evaluation personnel (Q15.6) was explainable by Q15.2 and Q15.5. Responses to “I trust the person evaluating me to be objective” (Q15.2) significantly predicted satisfaction with evaluation

personnel (Q15.6), $B = 0.56$, $t(51) = 4.98$, $p < .001$. Responses to “my evaluator is experienced with NTSG” (Q15.5) significantly predicted satisfaction with evaluation personnel, Q15.6, $B = 0.40$, $t(51) = 4.46$, $p < .001$. Table 25 summarizes the results of the regression model.

Table 25. Results for Linear Regression with Q15.2 and Q15.5 predicting Q15.6

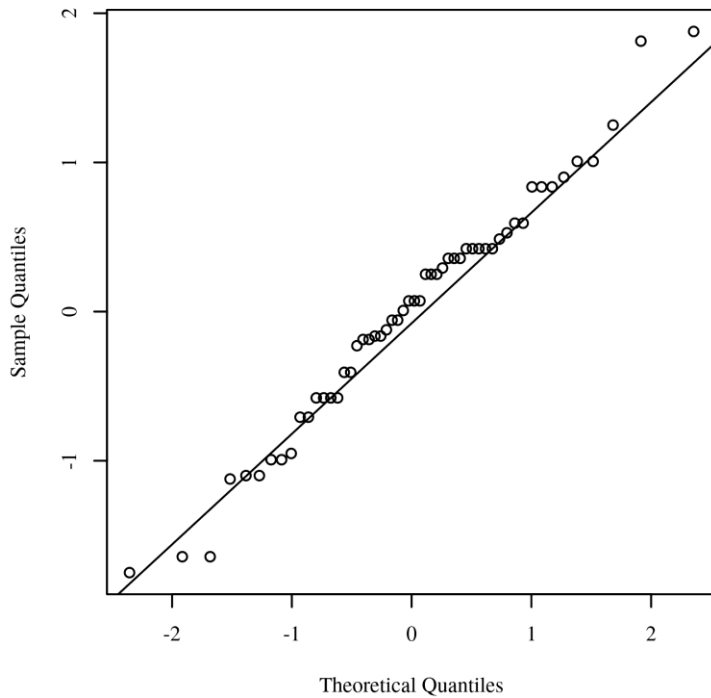
Variable	B	SE	95.00% CI	β	t	p
(Intercept)	-0.0009	0.40	[-0.80, 0.80]	0.00	-0.002	.998
Q15.2	0.56	0.11	[0.33, 0.79]	0.49	4.98	< .001
Q15.5	0.40	0.09	[0.22, 0.57]	0.44	4.46	< .001

Teacher Evaluation Product

A linear regression analysis was conducted and the results were that “my evaluation provides an accurate summary of my performance” (Q18.1) and “my evaluation has led to improvements in student learning” (Q18.6) significantly predicted satisfaction with the product of evaluation (Q18.7).

The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution. Figure 3 presents a Q-Q scatterplot of the model residuals.

Figure 3. Q-Q scatterplot for normality of the residuals for the regression model



Since the linear regression considered more than one independent variable, Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. All predictors in the regression model had VIFs less than 10. Table 26 presents the VIF for each predictor in the model.

Table 26. Variance Inflation Factors for Q18.1 and Q18.6

Variable	VIF
Q18.1	1.41
Q18.6	1.41

The results of the linear regression model were significant, $F(2,51) = 37.05, p < .001, R^2 = .59$, indicating that approximately 59.23% of the variance in satisfaction with the product of evaluation (Q18.7) was explainable by Q18.1 and Q18.6. “My evaluation provides an accurate summary of my performance” (Q18.1) significantly predicted satisfaction with the product of

evaluation (Q18.7), $B = 0.52$, $t(51) = 4.14$, $p < .001$. “My evaluation has led to improvements in student learning” (Q18.6) significantly predicted Q18.7, $B = 0.41$, $t(51) = 4.12$, $p < .001$. Table 27 summarizes the results of the regression model.

Table 27. Results for Linear Regression with Q18.1 and Q18.6 predicting Q18.7

Variable	<i>B</i>	<i>SE</i>	95.00% CI	β	<i>t</i>	<i>p</i>
(Intercept)	0.25	0.39	[-0.54, 1.04]	0.00	0.64	.524
Q18.1	0.52	0.13	[0.27, 0.77]	0.44	4.14	< .001
Q18.6	0.41	0.10	[0.21, 0.62]	0.44	4.12	< .001

Summary of Results

In Chapter IV, the researcher analyzed the demographic information collected from of high school music teachers in Virginia ($n = 76$) via the *Teacher Evaluation Satisfaction Survey (TESS)*. The levels of satisfaction with their evaluation (RQ 1) were reported and the results further identified factors related to their levels of satisfaction (RQ 2).

Demographic data included personal demographics – age, sex, and Race/Ethnicity; professional demographics – highest level of education and years of service; and teaching context – school type, school size, school setting, subject taught, and evaluation personnel.

Research Question 1

In determining the levels of satisfaction with the teacher evaluation, three prompts related to levels of satisfaction with the process, personnel, and product of teacher evaluation were used. The mean reported levels of satisfaction with “the process of my evaluation” was $M = 3.5$. Over half the participants somewhat agreed ($n = 14$, 25.0%) or strongly agreed ($n = 15$, 26.8%) responding to the prompt, “I am satisfied with the process of my evaluation.” The mean reported levels of satisfaction with “the personnel who conducted my evaluation” was $M = 3.4$. Slightly

over half the participants somewhat agreed ($n = 13, 23.6\%$) or strongly agreed ($n = 15, 27.3\%$) responding to the prompt, “I am satisfied with the personnel who conduct my evaluations.” The mean reported levels of satisfaction with “the product of my evaluation” was $M = 3.1$. Less than half the participants somewhat agreed ($n = 15, 28.8\%$) or strongly agreed ($n = 8, 14.3\%$) responding to the prompt, “I am satisfied with the product of my evaluation.”

A series of one-way analyses of variance (ANOVA) were conducted to determine whether there were significant differences in satisfaction with teacher evaluation product by category: age, sex, Race/ethnicity, highest level of education, years of service, school type, school size, school setting, or the person performing the evaluation. The ANOVA were examined based on an alpha value of .05. The results of the ANOVA were not significant, $F(21, 30) = 1.29, p = .257$, indicating the differences in satisfaction with the teacher evaluation product were similar across all demographic categories.

Averaging the three categories of satisfaction (means of the means) to determine a grand mean to serve as the overall levels of satisfaction with teacher evaluation (RQ1) produces a result of $M = 3.3$ where just over half of the participants report satisfaction with their teacher evaluation. A Levels of satisfaction Composite scale was calculated consisting of the three previously mentioned categories: teacher evaluation process, teacher evaluation personnel, and teacher evaluation product. A Cronbach alpha coefficient of .85, indicated good reliability and consistency between scores.

Research Question 2

Multiple linear regressions were conducted to identified predictors related to their levels of satisfaction. When evaluating the levels of satisfaction with the process of evaluation, the results were significant, $F(2,52) = 60.82, p < .001, R^2 = .70$, indicating that approximately

70.05% of the variance in satisfaction with the process of evaluation was explainable by responses to “my evaluation highlights teacher practices” and “the criteria used are appropriate for HS music teacher evaluation.”

When evaluating the levels of satisfaction with evaluation personnel, the results of the linear regression model were also significant, $F(2,51) = 53.17, p < .001, R^2 = .68$, indicating that approximately 67.59% of the variance in satisfaction with the evaluation personnel was explainable by responses to “I trust the person evaluating me to be objective” and “my evaluator is experienced with NTSG.”

When evaluating the levels of satisfaction with the product of evaluation, the results of the linear regression model were once again significant, $F(2,51) = 37.05, p < .001, R^2 = .59$, indicating that approximately 59.23% of the variance in satisfaction with the product of evaluation was explainable by responses to “My evaluation provides an accurate summary of my performance” and “My evaluation has led to improvements in student learning.”

CHAPTER V: DISCUSSION

The purpose of this study was to investigate high school music teachers' levels of satisfaction with the evaluative practices used by administrators and music supervisors in the Commonwealth of Virginia. Virginia high school music teachers ($n = 76$) were surveyed to collect data to determine their levels of satisfaction with their evaluation (RQ 1) and what, if any, factors related to their levels of satisfaction (RQ 2). The satisfaction level was measured using a researcher-constructed *Teacher Evaluation Satisfaction Survey (TESS)* completed by high school music teachers in Virginia. The data collected were used to answer the following research questions:

1. What are the levels of satisfaction, among high school music teachers, using contemporary evaluative practices?
2. Are there specific factors that relate to the levels of satisfaction with teacher evaluation?

Participants were (a) high school teachers, (b) music teachers, and (c) in the Commonwealth of Virginia and were identified through their membership in the Virginia Music Educators Association (VMEA). Participants were invited to participate in the study through the VMEA Newsletter and mailing lists.

To measure the levels of satisfaction with current evaluative practices, a survey was constructed to collect data on music teacher perceptions on the levels of satisfaction with the methods used in their evaluation. The researcher constructed *Teacher Evaluation Satisfaction Survey (TESS)* containing three parts (Appendix B): Part 1 collected demographic data on the participants, Part 2 used a *Likert* scale for participants' responses on their perceptions regarding the *process*, *personnel*, and *product* of their evaluation, and Part 3 contained open-ended or free-

response questions allowing participants to share their lived-experiences in their own words. The researcher utilized a focus group of higher education music education professionals and a pilot group of high school music teachers in Virginia before distributing the final version of the survey.

Research Question 1

“What are the levels of satisfaction, among high school music teachers, using contemporary evaluative practices?” In determining the levels of satisfaction with how high school music teachers were evaluated, RQ1 assessed the levels of satisfaction in three areas: the process, personnel, and product of teacher evaluation. A summary is shown in Table 28.

Table 28. Reported levels of satisfaction with evaluation

Variable	<i>M</i>	<i>SD</i>	<i>n</i>	Strongly Disagree (%)	Some-what disagree (%)	Neither agree nor disagree (%)	Some-what Agree (%)	Strongly agree
I am satisfied with the process of my evaluation	3.5	1.2	56	<i>n</i> = 5 (8.9%)	<i>n</i> = 4 (7.1%)	<i>n</i> = 18 (32.1%)	<i>n</i> = 14 (25.0%)	<i>n</i> = 15 (26.8%)
I am satisfied with the personnel who conduct my evaluations	3.4	1.3	55	<i>n</i> = 6 (10.9%)	<i>n</i> = 8 (14.6%)	<i>n</i> = 13 (23.6%)	<i>n</i> = 13 (23.6%)	<i>n</i> = 15 (27.3%)
I am satisfied with the product of my evaluation	3.1	1.3	56	<i>n</i> = 9 (16.1%)	<i>n</i> = 5 (8.9%)	<i>n</i> = 19 (33.9%)	<i>n</i> = 15 (26.8%)	<i>n</i> = 8 (14.3%)

Fifty-six participants answered these three prompts. The grand mean of the three areas measuring the levels of satisfaction was $M = 3.3$, which seems to indicate that although just over half of the participants were satisfied, there is much room for improvement. The mean score ($M = 3.3$) was closest to the “neither agree nor disagree” rating (3) on the Likert scale. A stronger

result in either direction (2 = somewhat disagree or 4 = somewhat agree) would be more conclusive.

Of the resources reviewed in Chapter II, many focused on the evaluation of teachers and the impact of federal legislation over the last twenty years. Many of the same topics were present: high stakes teacher evaluation (HSTE), evaluation of teachers in non-tested grades and subjects (NTGS), use of value-added measures (VAMs), and formative and summative teacher evaluation. The large quantity of journal articles, dissertations, books, and studies from across the country covering all teaching subjects reveals that teachers, administrators, lawmakers, and parents are interested in teacher evaluation. The importance of the current study lies in answering the first research question.

Teachers participating in this study reported levels of satisfaction with “the process of my evaluation” was $M = 3.5$. Is this good? How does this response compare to other states or national measures? The National Center for Education Statistics reported the results of the 2017-2018 National Teacher and Principal Survey. The results were that 86% of high school teachers reported “overall, the evaluation process was fair” which is higher than the 52% of participants who somewhat or strongly agreed with “I am satisfied with the process of my evaluation.” Further research is needed to determine what factors are related to the difference in these responses.

When responding to “the results of my evaluation were accurate,” the NCES survey responses showed 86% agreement. This prompt corresponds to the *TESS* prompt regarding satisfaction with the *product* of evaluation where only 43% somewhat or strongly agreed that they were satisfied with the results or product of their evaluation.

These two comparisons indicate a sizeable difference between national high school teacher perceptions regarding their evaluation and Virginia high school music teachers' responses in the current study. Can the differences be attributed to geographic differences? Are the differences related to subject area? The literature reviewed in this study showed how broad the scope is and how many different factors contribute to perceptions of satisfaction with teacher evaluation.

The data presented in Chapter IV: Results showed that the answer to RQ 1 is $M = 3.5$ (process), $M = 3.4$ (personnel), and $M = 3.1$ (product). Although these are technically answers to the research question, the answer presents a new list of questions. Some of these questions are why are the levels of satisfaction what they are? Should they be closer to national averages? RQ 2 moved in the direction of providing some additional information.

Research Question 2

“Are there specific factors that relate to the levels of satisfaction with teacher evaluation?” Before analyzing the data to answer RQ2, consideration was given to the possibility of various factors' effects on satisfaction with teacher evaluation. A Kruskal-Wallis rank sum test was conducted to assess if there were significant differences in satisfaction with evaluation between the levels of various demographic data. For each area – process, personnel, and product – the results were not significant based on an alpha value of .05, indicating that the mean rank of satisfaction with evaluation was similar for each level of these demographic characteristics. Factors possessing a predictive relationship with levels of satisfaction were identified through linear regression analyses.

Teacher Evaluation Process

For satisfaction with the process of evaluation, “my evaluation highlights teacher practices” and “the criteria used are appropriate for HS music teacher evaluation” explained approximately 70% of the variance in satisfaction with the process of evaluation. Participants who believed that what they did as teachers (teacher practices) and that their evaluation criteria were appropriate showed a measurable difference in levels of satisfaction with the process of their evaluation. Literature reviewed in Chapter II showed similar findings. Concerns over the inappropriate use of VAMs in the evaluation of teachers in NTGSs were among the most commonly reported (Gagnon et al., 2016; Phillips et al., 2017).

Teacher Evaluation Personnel

Regarding satisfaction with the personnel conducting the evaluation, “I trust the person evaluating me to be objective” and “my evaluator has experience with NTGS” explained approximately 67.6% of the variance in satisfaction with the evaluation personnel. Participants who indicated a level of trust with their evaluator and believed their evaluator had experience with non-tested grades and subjects showed a measurable difference in level of satisfaction with the evaluation personnel. These relational characteristics of the evaluator were also present in the literature reviewed. The Emotional Intelligence (EI) of administrators (Todd, 2022) and level of trust in the personnel conducting the evaluation (Harris, 2018) were reported as relevant factors in teacher evaluation and job satisfaction.

Evaluators’ familiarity with NTGS featured prominently in studies (Papay, 2012; Salvador & Krum, 2019), dissertations (Harris, 2018; Kennedy, 2021), and position statements by professional organizations like the National Association for Music Education (NAfME). The high level of concern for improper evaluation of teachers in NTGSs (Gagnon et al., 2016;

Phillips et al., 2017) was echoed by the predictive relationship between the level of satisfaction with the evaluating personnel and their experience with NTGS.

Teacher Evaluation Product

For satisfaction with the product of evaluation, “my evaluation provides an accurate summary of my performance” and “my evaluation has led to improvements in student learning” explained approximately 59.2% of the variance in satisfaction with the product of evaluation. Not surprisingly, participants who believed that their evaluation accurately summarized their performance and led to improvements in student learning showed a measurable difference in levels of satisfaction with the product of their evaluation.

The accurate summary of teacher performance and improvement in student learning align the two most foundational elements of teaching and evaluation: teacher practices and student learning. Jenkins (2018) addressed teacher practices in his definition of teacher evaluation as “a systematic, ongoing process used to assess teachers’ competence, performance, and effectiveness in the classroom” (p. 1658). The evaluation of teacher performance is reflected in evaluation rubrics and was the subject of *The Widget Effect* (Weisberg et al., 2009) where the impact of individual teacher performance or practices are minimized by the assumption that teacher effectiveness is the same and judged as a whole.

In Their Own Words

Participant response to the open-ended items provided insight on what factors they believed related to levels of satisfaction with evaluation. Their opinions and perspectives are the foundation for their responses to the items on this survey. When prompted to list “factors, practices, or characteristic that have a positive impact on your levels of satisfaction with music teacher evaluation,” participants cited both factors for themselves and factors for their evaluators.

Participants valued the opportunity to have input into their own evaluation. This included teacher-provided measures of student learning in music and descriptions of what teacher effectiveness looked like in a music classroom which is consistent with research on the relationship regarding the administrators/teacher relationship (Opper, 2019; Price, 2012). Participants also expressed their opinions on evaluator qualities that had a positive effect including, evaluators with content-knowledge in music and evaluators who were respectful, supportive, and showed a personal interest in the success of those they were evaluating. This is consistent with the literature reviewed in Chapter II (Harris, 2018; Todd, 2022).

When responding to the prompt, “Factors, practices, or characteristic that have a negative impact on your levels of satisfaction with music teacher evaluation,” the participants’ responses were, not surprisingly, the opposite of those factors having a positive impact. They also cited use of rubrics that don’t “fit” music teacher evaluation and lack of communication as negative factors which is consistent with the findings presented in the Literature Review (Gerruty, 2013; Taebel, 1990).

Possible Limitations

Generalization of the results of the current study to the larger population should be considered with caution due to the sample size ($n = 76$). Survey administration method may have had an impact on response rates as the number of participants completing the survey ($n = 56$) differed from the number of participants starting the survey ($n = 76$). The responses from those completing only the demographic portion showed survey progress of 52% and were reported in Part 1, but not in Parts 2 and 3 of the survey. It is not clear to the researcher if this was a Qualtrics glitch or some other technical or distribution malfunction.

Response rate determination was also hampered due to the nature of the invitations. The researcher was unable to determine the number of VMEA members who saw the invitation and chose not to participate. Another relevant piece of data that was unobtainable was the number of VMEA members by grade level. VMEA does not categorize its 3,200 members by grade level taught.

Implications

Although there were some positive indicators in the levels of satisfaction with evaluation expressed by the data, better communication between teachers and evaluators could lead to improvements on multiple levels including, teacher satisfaction with evaluation, evaluator ability to assess areas outside content areas of experience, improved teacher effectiveness, and improved student learning.

Among the suggestions for increased satisfaction with evaluation, teachers listed the ability to set student growth goals and the methods by which the growth would be determined. Teachers also indicated that taking the initiative in pre-observation conferences and showing the evaluator what they are looking for within the context of the music classroom had a positive impact on their evaluation. The implication here is that the lack of experience of an evaluator becomes an opportunity for the music teacher (content expert) to demonstrate teaching proficiency and effectiveness in their context and show the evaluator how those practices “fit” into the evaluation rubric being used.

When describing characteristics of a “good” evaluator, participants listed (a) understanding, objective, willing to learn, (b) content knowledge, and (c) good communication skills. This list was similar to the “advice you would give to new music teachers” regarding their

evaluation and implied an increased level of agency or self-advocacy in the evaluation process which is analogous to lighting a candle rather than cursing the darkness.

In contrast to the statements focused on content area knowledge, other participant responses echoed the “teaching is teaching” mantra where evaluation identifies good practices superseding content area. Teacher effectiveness and teacher practices can be improved via evaluation by administrators with or without content knowledge. Study participants listed another characteristic of a good evaluator – “willing to take the time and ask questions when they don't understand something” – which provides the teacher an opportunity to be the content expert. Conversely, the teacher can also learn from the experience and expertise of the evaluator in practices outside the content area.

Another qualitative aspect of the study indicates an attitudinal component to satisfaction with teacher evaluation. Participants responded based on their previous experiences with evaluation. Some indicated negative experiences and resulting pessimistic outlooks and attitudes regarding teacher evaluation. Others showed more optimistic attitudes, possibly linked to positive experiences. Although there may be a correlation between attitude and experience, it was beyond the scope of the current study but could be investigated in future research.

A frequent theme in the open-ended items was the desire for evaluators with content-knowledge in music. The demonstration of teaching effectiveness and best practices in the music classroom can be very different from traditional classrooms. Evaluators with content-knowledge in music were mentioned multiple times ($n = 14$) in the open-ended item responses and content knowledge is a foundational component in the various position statements issued by NAFME (Appendix A). The surprise came in regarding participant responses to “my evaluator is competent to evaluate my teaching in my context” and “my evaluator has experience in music

supervision” which one would assume would be predictive of the levels of satisfaction with evaluation personnel. Linear regression did not indicate a statistically significant predictive relationship between these factors and levels of satisfaction with evaluation personnel. Instead, the levels of satisfaction were related to “trust” with the evaluator’s objectivity and experience with NTGS evaluation.

A related observation was that numerous open-ended item responses called for evaluation by music specialists. The data showed personnel conducting the evaluations were overwhelmingly “building administrators” ($n = 69, 92\%$) and not music supervisors. It is notable that the levels of satisfaction with evaluation personnel ($M = 3.4$) and over half of the participants reported “somewhat agree” ($n = 13, 23.6\%$) or “strongly agree” ($n = 15, 27.3\%$) regarding their satisfaction with the personnel conducting their evaluations. The implication is that the perceived need for music specialists for music teacher evaluation is not supported by the data collected by the current survey and that participants have experienced some levels of satisfaction ($M = 3.3$) with their evaluations performed overwhelmingly by building administrators.

Additionally, knowledge of the evaluation process and criteria did not directly relate to levels of satisfaction with the process. Linear regression analysis did not confirm a predictive relationship. Instead, there was a significant relationship ($F(2,52) = 60.82, p < .001, R^2 = .70$) between use of appropriate criteria for music teacher evaluation that highlighted teacher practices and levels of satisfaction with the process of their evaluation.

The implication of generalizability is supported in the similarity in national high school teacher demographic characteristics and the participants in this study. Similarity in demographic

categories strengthens the case for generalizability of findings in research. See Table 29 for a comparison of *TESS* results and National High School Teacher demographics.

Table 29. *TESS* Results to National HS Teachers

Category	TESS Participants	National HS Teachers
Age	58% were 31-50 years-old	56.4% were 30-49 years-old
Sex	49% female 48% male	60% female 40% male
Race/Ethnicity	80% White	80% White
	10% Hispanic	8% Hispanic
	5% African-American	6% African-American
Level of Education	55% Master’s Degree	53% Master’s Degree
Years of Service	44.5% 15 years or more	44% 15 years or more
School Type	88% Public 11% Private	90% Public 10% Private

Note. 2017–18 National Teacher and Principal Survey (<https://nces.ed.gov/pubs2020/2020142rev.pdf>)

Suggestions for Future Research

Future research could expand the scope of the current study to include more participants to make the sample size appropriate for generalization to the larger population of high school music teachers. The scope could also be broadened to “music teachers” to accommodate middle and elementary school music teachers. As the geographic scope of this study was limited to Virginia, the study could be replicated in other states or nationally for comparison.

In addition to the broadening mentioned above, the focus of study could also be narrowed to investigate differences between grade levels and even differences among specific music-subject teachers (general music, choir, band, etc.). Numerous studies have considered the differences between school size and setting. These could also be investigated in future studies to determine if the levels of satisfaction with evaluation is impacted.

The results of the current study indicate that the relationship between the teacher and evaluator may impact the levels of satisfaction. In addition, evaluator characteristics and

dispositions mentioned in the open-ended items were frequently cited by the survey participants. Future research on organizational and interpersonal dynamics within the school environment, as well as characteristics of building administrators and their relationship to satisfaction with evaluation, could provide insight into how to improve satisfaction with evaluation (as experienced by administrator and teacher), teacher effectiveness, and student learning.

Research regarding characteristics of a “successful” administrator/evaluator might also produce data that could be applied to music and other non-tested grades and subjects. This could also provide professional development opportunities for administrators lacking experience in evaluation of NTGS teachers (Kennedy, 2021).

One study participant suggested further research reporting on music teacher evaluation results. Future research with specific evaluation outcomes could provide more data on how to increase the levels of satisfaction with evaluation, improve teacher effectiveness, and student learning. Professional development opportunities for music teachers and others could be suggested based on this further research.

Recommendations

Based on the findings of the current study, the researcher recommends continued conversations, presentations, professional development, and the development of new and innovative evaluation tools and practices for the successful and accurate evaluation of high school music teachers. Campus-based or district wide emphasis on collaboration between administrators and music teachers regarding teacher evaluation should increase. Teachers in NTGS, including fine arts and music, should be proactive in offering solutions to problems posed by ill-fitting rubrics. Evaluators should be transparent with music teachers on the difficulty of evaluating music teachers. This collaboration could include focus group discussion between

department chairs, fine arts specialists, and building administrators where challenges and solutions arising in music teacher evaluation can be addressed. Professional development opportunities presented by administrators and master teachers on the subject or teacher evaluation should be offered.

Teacher evaluation systems should strive for balance between formative and summative goals with a focus on teacher growth and improvement for the purpose of student growth (Ford & Hewett, 2020). Utilizing existing or newly created integrated frameworks, teacher evaluation should and could consider the varied but mutual interests of both evaluators and teachers. Echoing the findings of Ford and Hewett (2020), evaluation systems should be formative and summative and consist of five key characteristics: (a) compatible goals (individual and organization), (b) two-way communication, (c) a supportive climate, (d) technical rationality, and (d) use of multiple data sources (p. 10).

Conclusion

The purpose of this study was to investigate high school music teachers' levels of satisfaction with the evaluative practices used by administrators and music supervisors in the Commonwealth of Virginia. The data collected were used to answer the following research questions:

1. What are the levels of satisfaction, among high school music teachers, using contemporary evaluative practices?
2. Are there specific factors that relate to the levels of satisfaction with teacher evaluation?

Results from the *Teacher Evaluation Satisfaction Survey (TESS)* were analyzed to determine levels of satisfaction with their evaluation and what, if any, factors related to their

levels of satisfaction. While participants reported a general “satisfaction” with their evaluation, the data revealed areas for improvement including both evaluator and teacher practices. Participants revealed an undertone of anxiety in their open-ended items also reflected in a newsletter from the National Education Policy Center, “Teachers are twice as stressed out as other American employees” (NEPC, 2022). Stress associated with teacher evaluation can be mitigated by providing accurate, valuable, reliable teaching evaluations for EVERY teacher. The evaluation process can help guide teachers, help them feel professionally stable and effective, and improve the quality of their teaching. Most importantly, teacher evaluation can help provide the best, most effective teachers for America’s children. Music is a core subject and it deserves teachers who are evaluated with the rigor consistent of each and every teacher in the school systems.

REFERENCES

- Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1), 54-76. Retrieved March 19, 2021, from <http://www.jstor.org/stable/44984570>
- Amrein-Beardsley, A., & Geiger, T. J. (2019). Potential sources of invalidity when using teacher value-added and principal observational estimates: artificial inflation, deflation, and conflation. *Educational Assessment, Evaluation and Accountability*, 31(4), 465-493. <http://dx.doi.org/10.1007/s11092-019-09311-w>
- Bernard, C. F. (2015). *Ensemble educators, administrators, and evaluation: support, survival, and navigating change in a high-stakes environment* (Order No. 3704455). Available from ProQuest Central; ProQuest Dissertations & Theses Global; Social Science Premium Collection. (1687150752). <https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/dissertations-theses/ensemble-educators-administrators-evaluation/docview/1687150752/se-2>
- Blumberg, A. (1985). Where we came from: Notes on supervision in the 1840's. *Journal of Curriculum and Supervision*, 1(1), 56-65.
- Brophy, T. (1993). Evaluation of music teachers: Toward defining an appropriate instrument. Unpublished paper. Retrieved from <http://0-files.eric.ed.gov.opac.msmc.edu/fulltext/ED375029.pdf>
- Close, K., & Amrein-Beardsley, A. (2018). Learning from what doesn't work in teacher evaluation. *Phi Delta Kappan*, 100(1), 15-19. doi:10.1177/0031721718797115
- Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician*, 35(3), 124-129. <https://doi.org/10.1080/00031305.1981.10479327>

- Creswell, J. (2014). *Research design: qualitative, quantitative, and mixed methods approaches* (Fourth edition.). SAGE Publications.
- Croft, Michelle, & Buddin, R. (2015). Applying value-added methods to teachers in untested grades and subjects. *Journal of Law and Education*, 44(1), 1-22.
<http://ezproxy.liberty.edu/login?url=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Fapplying-value-added-methods-teachers-untested%2Fdocview%2F1644935556%2Fse-2%3Faccountid%3D12085>
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292-307. <https://doi.org/10.1037/1082-989X.2.3.292>
- Dee, T., & Wyckoff, J. (2017). A lasting impact: High-stakes teacher evaluations drive student success in Washington, D.C. *Education Next: A Journal of Opinion and Research.*, 17(4).
- DeSoto, D. A. (2018). *Relation of High Stakes Teacher Evaluation Implementation in Hawaii to Teacher Satisfaction* Available from Social Science Premium Collection. (2130846717; ED586889).
<https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/dissertations-theses/relation-high-stakes-teacher-evaluation/docview/2130846717/se-2>
- Donaldson, M. L., Papay, J. P. (2015). Teacher evaluation for accountability and development. In Ladd, H. F., Goertz, M. E. (Eds.), *Handbook of research in education finance and policy* (pp. 174–193). New York, NY: Routledge.
- Emert, D., Sheehan, S., & Deitz, O. D. (2013). Measuring music education: Music teacher evaluation in Pennsylvania. *Music Educators Journal*, 100(1), 30-31.
[doi:10.1177/0027432113495796](https://doi.org/10.1177/0027432113495796)

- Finn, C. (2015). Transforming tomorrow's schools: The wheels of education reform grind exceedingly slow, but they have ground out some progress. We need more. *Hoover Digest., 1*.
- Ford, T. G. (2018). Pointing teachers in the wrong direction: Understanding Louisiana elementary teachers' use of compass high-stakes teacher evaluation data. *Educational Assessment, Evaluation and Accountability, 30*(3), 251-283. doi:10.1007/s11092-018-9280-x
- Ford, T. G., & Hewitt, K. (2020). Better integrating summative and formative goals in the design of next generation teacher evaluation systems. *Education Policy Analysis Archives, 28*(63). <https://doi.org/10.14507/epaa.28.5024>
- Froehlich, H., Frierson-Campbell, C., Rohwer, D., McCarthy, M., & Coan, D. (2013). *Inquiry in music education: Concepts and methods for the beginning researcher*. Routledge.
- Gall, M., Gall, J., & Borg, W. (2007). *Educational research: An introduction* (Eighth edition.). Pearson/Allyn & Bacon.
- Gagnon, D. J., Hall, E. L., & Marion, S. (2016). Teacher evaluation and local control in the US: An investigation into the degree of local control afforded to districts in defining evaluation procedures for teachers in non-tested subjects and grades. *Assessment in Education: Principles, Policy & Practice, 24*(4), 489-505. doi:10.1080/0969594x.2016.1167669
- Gerrity, K. W. (2013). Measuring music education: Teacher evaluation in Indiana. *Music Educators Journal, 99*(4), 17-19. doi:10.1177/0027432113486414

- Gilbert, A. D. (2016). The framework for 21st century learning: A first-rate foundation for music education assessment and teacher evaluation. *Arts Education Policy Review*, 117(1), 13-18. doi:10.1080/10632913.2014.966285
- Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on student test scores. *Educational Testing Service*. retrieved from <https://www.ets.org/Media/Research/pdf/PICANG14.pdf>
- Harris, D. N., & Herrington, C. D. (2015). Editors' introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, 44(2), 71–76. <https://doi.org/10.3102/0013189X15576142>
- Harris, J. Q., Sr. (2018). *Improving Music Teacher Effectiveness: An Examination of Resources and Strategies Used to Cultivate the Growth of Music Teachers* (Order No. 10784891). Available from ProQuest Central; ProQuest Dissertations & Theses Global; Social Science Premium Collection. (2038379267). <https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/dissertations-theses/improving-music-teacher-effectiveness-examination/docview/2038379267/se-2>
- Hewitt, K., & Amrein-Beardsley, A. (2016). *Student growth measures in policy and practice intended and unintended consequences of high-stakes teacher evaluations*. Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-53901-4>
- Hirokawa, J. O. (2013). *Evaluating Music Teachers: A Comparison of Evaluations by Observers with Varied Levels of Musical and Observational Background* Available from Social Science Premium Collection. (1697492388; ED555907). <https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/dissertations-theses/evaluating-music-teachers-comparison-evaluations/docview/1697492388/se-2>

- Jenkins, P. (2018). Teacher evaluation. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vol. 1, pp. 1658-1658). SAGE Publications, Inc., <https://www.doi.org/10.4135/9781506326139.n687>
- Jones, S. D., & Workman, E. (2016). ESSA's well-rounded education. *Education Commission of the States*. http://www.ecs.org/wp-content/uploads/ESSAs-Well-Rounded-Education-1.pdf?utm_source=Education+News+--+July+5%2C+2016&utm_campaign=Education+News+from+NNSTOY&utm_medium=email
- Katz-Cote, H. (2016). *A matrix of music supervisors' stories in the midst of school reform* (Order No. 10135058). Available from ProQuest Central; ProQuest Dissertations & Theses Global; Social Science Premium Collection. (1802533373). <https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/dissertations-theses/matrix-music-supervisors-stories-midst-school/docview/1802533373/se-2>
- Kennedy, Thomas Carl, I.,II. (2021). *Texas Administrators' Perspective of Music Educators in Classroom Evaluations in Regards to T-TESS* (Order No. 28869114). Available from Dissertations & Theses @ Liberty University; ProQuest Dissertations & Theses Global. (2620738001). <https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/dissertations-theses/texas-administrators-perspective-music-educators/docview/2620738001/se-2>
- Lavigne, A. L. (2014). Exploring the intended and unintended consequences of high-stakes teacher evaluation on schools, teachers, and students. *Teachers College Record*, 116(1), 1-29.

- Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010). *Methods in educational research: From theory to practice*. ProQuest Ebook Central <https://ebookcentral-proquest-com.ezproxy.liberty.edu>
- Malunda, P., Onen, D., Musaaazi, J. C. S., & Oonyu, J. (2016). Teacher evaluation and quality of pedagogical practices. *International Journal of Learning, Teaching, and Educational Research*, 15(9), 118-133. <https://www.ijlter.org/index.php/ijlter/article/view/753/pdf>
- Marshall, K. (2013). *Rethinking teacher supervision and evaluation how to work smart, build collaboration, and close the achievement gap* (2nd ed.). Jossey-Bass.
- Marzano, R., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching*. ProQuest Ebook Central <https://ebookcentral-proquest-com.ezproxy.liberty.edu>
- Marzano, R., Pickering, D., & Pollock, J. (2001). *Classroom instruction that works: research-based strategies for increasing student achievement*. Association for Supervision and Curriculum Development.
- Menard, S. (2009). *Logistic regression: From introductory to advanced concepts and applications*. Sage Publications. <https://doi.org/10.4135/9781483348964>
- National Education Association. (2010). Teacher assessment and evaluation: The National Education Association's framework for transforming education systems to support effective teaching and improve student learning [White paper]. <https://files.eric.ed.gov/fulltext/ED583104.pdf>
- National Education Policy Center [NEPC]. (2019, July 11). *Newsletter: This teacher evaluation trend was the reform du jour. Now states are backing away*. Author. Retrieved from: <https://nepc.colorado.edu/sites/default/files/publications/Newsletter%20teacher%20eval.pdf>

- Nielsen, L. (2014). Teacher evaluation: Archiving teaching effectiveness. *Music Educators Journal*, 101(1), 63-69. Retrieved March 12, 2021, from <http://www.jstor.org/stable/43289093>
- Orzolek, D. (2014). Measuring music education: Teacher evaluation in Minnesota. *Music Educators Journal*, 100(4), 22–24. <https://doi.org/10.1177/0027432114522387>
- Overland, C. T. (2014). Teacher evaluation and music education: Joining the national discussion. *Music Educators Journal*, 101(1), 56–62. <https://doi.org/10.1177/0027432114534448>
- Opper, I. (2019). Teachers Matter: Understanding Teachers' Impact on Student Achievement. Santa Monica, CA: RAND Corporation. https://www.rand.org/pubs/research_reports/RR4312.html.
- Papay, J.P. (2012). Refocusing the debate: Assessing the purpose and tools of teacher evaluation. *Harvard Educational Review*, 82 (1), 123-141.
- Parkes, K., Rohwer, D., & Davison, D. (2015). Measuring student music growth with blind-reviewed portfolios: A pilot study. *Bulletin of the Council for Research in Music Education*, (203), 23-44. doi:10.5406/bulcouresmusedu.203.0023
- Phillips, S. R., Mercier, K., & Doolittle, S. (2017). Experiences of teacher evaluation systems on high school physical education programs. *Physical Education and Sport Pedagogy*, 22:4, 364-377. DOI: 10.1080/17408989.2016.1241224
- Potter, D. (2021). Music teacher perceptions of evaluation: A case study of the Tennessee fine arts portfolio model. *Arts Education Policy Review*, ahead-of-print(ahead-of-print), 1-13. <https://doi.org/10.1080/10632913.2021.1992324>

- Price, H. E. (2012). Principal–Teacher Interactions: How Affective Relationships Shape Principal and Teacher Attitudes. *Educational Administration Quarterly*, 48(1), 39–85.
<https://doi.org/10.1177/0013161X11417126>
- Robelen, E. W. (2013). Classroom Portfolios Used as Alternative Teacher-Evaluation Measure. *Education Week*, 33(04), 1.
https://link.gale.com/apps/doc/A344250049/BIC?u=vic_liberty&sid=summon&xid=7ef319aa
- Robinson, M. (2015). The inchworm and the nightingale: On the (mis)use of data in music teacher evaluation. *Arts Education Policy Review*, 116(1), 9-21. doi: 10.1080/10632913.2014.944966
- Robinson, M. (2017). Music teachers' perceptions of high stakes teacher evaluation. *Arts Education Policy Review*, 120(1), 45-56. doi:10.1080/10632913.2017.1373380
- Shaw, R. (2019). Jumping through hoops: Troubling music teacher dilemmas in the new era of accountability. *Music Educators Journal*, 105(3), 23–29.
<https://doi.org/10.1177/0027432118816146>
- Shuler, S. C. (2012). Music education for life: Music assessment, part 2—Instructional improvement and teacher evaluation. *Music Educators Journal*, 98(3), 7-10.
doi:10.1177/0027432112439000
- Steinberg, M. P., Donaldson, M. L. (2016). The new educational accountability: Understanding the landscape of teacher evaluation in the post NCLB era. *Education Finance and Policy*, 11(3), 340–359.
- Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of*

Personnel Evaluation in Education, 20(3-4), 165-184. <http://dx.doi.org/10.1007/s11092-008-9053-z>

Taebel, D. (1990). Is evaluation fair to music educators? *Music Educators Journal*, 76(6), 50-54. Retrieved March 12, 2021, from <http://www.jstor.org/stable/3400968>

Todd, B. J. (2022). *A Case Study of the Emotional Intelligence Levels of a Secondary Principal and the Impacts upon Teacher Job Satisfaction and Organizational Commitment* (Order No. 29328136). Available from ProQuest Central; ProQuest Dissertations & Theses Global. (2703358369).
<https://go.openathens.net/redirector/liberty.edu?url=https://www.proquest.com/dissertations-theses/case-study-emotional-intelligence-levels/docview/2703358369/se-2>

U.S. Department of Education. (2004). *New no child left behind flexibility: Highly qualified teachers* [Web page]. Retrieved from <https://www2.ed.gov/nclb/methods/teachers/hqtflexibility.html>

U.S. Department of Education. (2012). *Race to the top fund* [Web page]. Retrieved from <http://www2.ed.gov/programs/racetothetop/index.html>

U.S. Department of Education. (2016). *Proposed ESSA regulation supports well-rounded education, protects all students* [Web page]. Retrieved from <https://www.ed.gov/news/press-releases/proposed-essa-regulation-supports-well-rounded-education-protects-all-students>

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009, 10). The widget effect. *The Education Digest*, 75, 31-35.
<http://ezproxy.liberty.edu/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fmagazines%2Fwidget-effect%2Fdocview%2F218196265%2Fse-2%3Faccountid%3D12085>

Zirkel, P. A., Green, & Oluwole, J. O. (2015). Adding value to teacher quality: All or nothing?/Response to "Applying value-added methods to teachers in untested grades and subjects". *Journal of Law and Education*, 44(3), 393-399,401-412.

<http://ezproxy.liberty.edu/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fscholarly-journals%2Fadding-value-teacher-quality-all-nothing-response%2Fdocview%2F1696203745%2Fse-2%3Faccountid%3D12085>

APPENDIX A: NAFME AND MENC POSITION STATEMENTS

1988 MENC Position Statement on Music Teacher Evaluation

MENC POLICY STATEMENT ON MUSIC TEACHER EVALUATION

Most states and school districts evaluate teachers for the purposes of improving instruction and to screen individuals who may not be effective in the classroom. The Music Educators National Conference (MENC) supports teacher assessment that is (1) designed to improve teaching competencies and (2) fair, reliable, and valid. However, the MENC also believes that the nature of music, which is both an aural art form and a performance art, requires music teachers to possess some special competencies in order to effect musical learning. Because of the unique nature of music and the special competencies required of music teachers, the MENC affirms the following:

- A. Existing evaluation programs need to be modified in some areas in order to adequately evaluate the performance competency of music teachers.
- B. Special assessment instruments or items are needed to evaluate the special competencies required of music educators.
- C. The evaluator or a member of the evaluation team should be knowledgeable in music.

-Adopted by the MENC National Executive Board at its April 1988 meeting

1993 MENC Criteria for Evaluation of Music Educators

The following is the list and the rater's instructions for evaluating the criteria:

Instructions: Relative to other music teachers you have known or worked with, how does this teacher compare with respect to each of the following characteristics? (Emphasis should be on teacher effectiveness in the classroom and the ability to create an environment conducive to student learning, not on personal traits.)

- is knowledgeable about music education and his or her field of specialization,
- demonstrates a high level of musicianship,
- plans learning experiences to achieve clearly defined musical goals,
- accurately diagnoses student learning problems and needs,
- communicates effectively with students,
- motivates students to achieve the highest possible level,
- uses appropriate and effective teaching materials,
- develops a supportive and stimulating environment characterized by a high level of student learning, excites students and interests them in music,
- emphasizes developing skills and knowledge to enable students to continue their music learning independently,
- pursues a systematic and effective personal program of professional growth and development,
- recognizes his or her weaknesses as a teacher and takes specific steps to remedy them,
- serves as a role model for students by exemplifying the musical and personal traits that the school seeks to develop, and
- contributes to the musical life of his or her community.

(MENC,1993)

NAfME

Teacher Evaluation (Position Statement)

The systematic application of student scores to teacher evaluation must be done carefully if the resulting systems for evaluation are truly to benefit our students and our schools. We urge all involved in the construction and implementation of these protocols and systems to carefully consider the importance of basing evaluation decisions on valid information. It is important for music educators and others involved in our schools to be aware of the following issues, to avert potential damage to school programs, teachers, and most of all, to students. To that end, the National Association for Music Education (NAfME) recommends the following:

1. Measures of student achievement used in teacher evaluation:

- Must be based on student achievement that is directly attributable to the individual teacher, in the subject area taught by that teacher. Student achievement measures must be used with care, ensuring that they accurately reflect a given teacher's contributions.
- Must be based on evaluation instruments that accurately reflect the achievements they purport to measure. This implies that the evaluation instruments are used by individuals with sufficient expertise to accurately observe and interpret the outcomes under measurement.
- Must be created to evaluate the curriculum that is taught. This implies that such measures reflect national, state, and local standards and curricula and use clear criteria known to the teacher in advance.
- Must be developed and applied in the context of the number of students taught and the instructional time available.
- Must take into account, if they are based on growth models, the beginning level of achievement from which growth is expected to take place. The evaluation instrument must be capable of capturing all levels of achievement, including the very highest levels of mastery.
- Must work on a multi-year cycle to allow for appropriate professional development and growth, enabling the evaluation to meet its primary goal of helping teachers to improve their service to students.

2. Successful Music Teacher Evaluation:

- Must include a balanced, comprehensive assessment of the teacher's contributions to student learning through multiple measures. These measures can and should collect information such as: (1) Indicators of teacher practice, such as planning and preparation (2) Indicators of the teacher's role in maintaining a productive classroom environment (3) Indicators that instruction is designed to reach specified goals (4) Indicators of teacher contribution to the school or district, as well as to the profession

of teaching at large (5) Indicators that students attain 21st century skills through instruction

- Must include measures of music student achievement along with the above indicators, as only one element of a teacher's evaluation. For evaluation of music teachers, measurements of student achievement should include evaluation in the three general areas of creating, performing, and responding. The relative weighting of measures in these three areas should be carefully designed to be commensurate with the nature of the class taught and the express educational goals for that class.
- Must, where the most easily observable outcomes of student learning in music are customarily measured in a collective manner (e.g., adjudicated ratings of large ensemble performances), limit the use of these data to valid and reliable measures and should form only part of a teacher's evaluation.
- Must avoid using school-wide measures other than those directly associated with music achievement. If the use of school-wide measures of attendance, dropout and graduation rates, and/or work habits is mandated, they should form a minimal part of the music teacher's evaluation.
- Must limit observation-based teacher evaluations to those conducted by individuals with adequate training in music as well as in evaluation.

APPENDIX B: TEACHER EVALUATION SATISFACTION SURVEY

A Descriptive Analysis of High School Music Teacher Evaluation in the Commonwealth of Virginia

Description of Study

This study is about high school teachers' levels of satisfaction with the evaluative measures used by administrators and music supervisors in the Commonwealth of Virginia. Teachers will be surveyed to gather data to determine if there is a significant difference between their respective levels of satisfaction and what, if any, factors create a measurable difference. The satisfaction level using this evaluation method will be measured using a researcher-constructed Teacher Evaluation Satisfaction Survey (TESS), which will be completed by high school music teachers. The data collected will be analyzed to determine what, if any modifications could be made to contemporary teacher evaluation practices.

Agreement to Participate

Project Title: A Descriptive Analysis of High School Music Teacher Evaluation in the Commonwealth of Virginia

Principal Investigator: Stephen Müller

How long will it take to complete this survey?

The survey should take 15-20 minutes to complete.

Will this study negatively affect me?

No. There are no risks of this study having a negative effect on you.

What do I get out of this research project?

You may help to develop an increased understanding of satisfaction with music teacher evaluation.

Will I get paid for participating?

There is no compensation.

What about my confidentiality?

We will do everything possible to make sure that your information is kept confidential. No identifiable information will be collected in the survey. Study data will be protected by a username and password. Study data will be accessed from within a secure network. Absolute

confidentiality of data provided through the Internet cannot be guaranteed due to the limited protections of Internet access. Please be sure to close your browser when finished so no one will be able to see what you have been doing.

What if I do not want to be in this research study?

You do not have to be part of this project. This project is voluntary. If you agree to participate, at any time in this project you may stop participating without penalty.

What if I have questions?

You can ask Stephen Müller (915-238-5275/swmuller@uncg.edu) or Dr. Patricia Sink (pesink@uncg.edu) anything about the study. If you have concerns about how you have been treated in this study call the Office of Research Integrity Director at 1-855-251-2351.

Teacher Evaluation Satisfaction Survey (TESS)

Part 1: Demographic Information

Personal

Q2 Age (select one):

- 20-30
- 31-40
- 41-50
- 51-60
- 61-70
- 71 or older
- Prefer not to say

Q8 Sex:

- Male
- Female
- Prefer not to say

Q9 Race/Ethnicity:

- White/Caucasian
- Asian – Eastern
- Asian – Indian
- Hispanic
- African-American
- Native-American
- Mixed race
- Other _____
- Prefer not to say

Professional

Q10 Highest level of education received:

- Bachelor's
- Master's
- Doctorate
- Other _____

Q12 Years of service:

- 0-5
- 6-15
- 15-25
- 25 or more

Teaching Context

- Q11 Subject taught (select all that apply):
Band
Choir
Orchestra
General Music
Theory
Piano
Guitar
Other _____
- Q3 School type:
Public
Private
Charter
- Q4 School size:
1,000 or less
1,001-2,000
2001 or more
- Q5 School setting:
Urban
Suburban
Rural
- Q7 Teacher Evaluations are performed by:
Building Administrator
Fine Arts Supervisor
Other _____
- Q14 Number of evaluations/year:
Less than 1
1-3
More than 3

Part 2: Process, Personnel, and Product of Teacher Evaluation

For the following statements, please select the response most accurately reflecting your agreement- (1) strongly disagree, (2) somewhat disagree, (3) neutral, (4) somewhat agree, (5) strongly agree.

Teacher Evaluation Process

- Q13.1 I understand the evaluation process for my position. 1 2 3 4 5
- Q13.2 I know the criteria upon which I am being evaluated. 1 2 3 4 5
- Q13.3 I am given time to prepare for my evaluation. 1 2 3 4 5
- Q13.4 My evaluation highlights teacher practices. 1 2 3 4 5
- Q13.5 My evaluation includes observation(s). 1 2 3 4 5
If so, how many per year? _____
- Q13.6 My evaluation includes student growth measures in my content area. 1 2 3 4 5
If so, what kind? _____
- Q13.7 My evaluation includes student growth measures outside my content area. 1 2 3 4 5
If so, what kind? _____
- Q13.8 The criteria used are appropriate for HS music teacher evaluation. 1 2 3 4 5
- Q13.9 I am satisfied with the process of evaluation at my school. 1 2 3 4 5

Q20 List some factors, practices, or characteristics that have a positive impact on your levels of satisfaction with music teacher evaluation.

Q21 List some factors, practices, or characteristics that have a negative impact on your levels of satisfaction with music teacher evaluation.

Teacher Evaluation Personnel

- Q15.1 I know the person who will be evaluating me. 1 2 3 4 5
- Q15.2 I trust the person evaluating me to be objective. 1 2 3 4 5
- Q15.2 My evaluator is competent to evaluate my teaching in my context. 1 2 3 4 5
- Q15.4 My evaluator has experience in music supervision. 1 2 3 4 5
- Q15.5 My evaluator is experienced with non-tested grades and subjects (NTGS). 1 2 3 4 5
- Q15.6 I am satisfied with the personnel who conduct my evaluations. 1 2 3 4 5

Q19 In your experience, what are some characteristics found in a “good” evaluator?

Teacher Evaluation Product

- Q18.1 My evaluation provides an accurate summary of my performance. 1 2 3 4 5
- Q18.2 My evaluation impacts my job security. 1 2 3 4 5
- Q18.3 My evaluation impacts advancement decisions. 1 2 3 4 5
- Q18.4 My evaluation impacts pay raise decisions. 1 2 3 4 5

- Q18.5 My evaluation has led to improvements in my teaching. 1 2 3 4 5
Q18.6 My evaluation has led to improvements in student learning. 1 2 3 4 5
Q18.7 I am satisfied with the product of my evaluation. 1 2 3 4 5

Part 3- Open-ended Items: In their own words

- Q22 What advice would you give to new music teachers to your campus regarding music teacher evaluations?
- Q23 What advice would you give to evaluators to your campus regarding music teacher evaluation?
- Q25 Is there anything you would like to share with me, as a researcher, that was not covered in this survey on music teacher evaluation?

APPENDIX C: OPEN-ENDED RESPONSES: IN THEIR OWN WORDS

Open-ended responses: In their own words

Q20 List some factors, practices, or characteristics that have a positive impact on your levels of satisfaction with music teacher evaluation.

Observations are very detailed and written down. We receive a follow up notification afterwards which details how and what we taught during our observation. Feedback is also given.

Generally benefit given to teacher if they follow a prescribed process of teaching.

None

that is know what it is and how to prepare for it

My administrator realizes that teaching/learning in band looks different than other content areas

I know what to do to make a non-musical admin happy

Getting feedback on positive areas of my classroom practice. Having administrators ask questions about things they don't know or understand.

not sure how to answer

During COVID everyone was pleased with the performance and willing to just support rather than evaluate and give feedback.

allowing me to write my own SMART goals to show student growth - this lets me use scales as a content area and I am not wasting time to check any boxes. The time I spend preparing can be meaningful and helpful to students

A quick review period with evaluator

Actual feedback on my instruction

The evaluation is focused on the project I choose to do for the year which will benefit my students and my own teaching.

Because I hardly ever see them, I am left alone to do what I need to do.

Administration comes to the room informally throughout the year.

I can set my own student progress goal.

When they stay for the whole class period

The understanding that music teachers and their classes are different than normal classrooms. The activities, student engagement, and assessments are all noticed and pointed out that they are different from a normal class, but appropriate in the music classroom.

Support from administration.

I appreciate when my observations are more than 15 minutes. I once had a principal who sat in my ensemble and sang with my students rather than standing in the back. I really appreciated her efforts to experience my class the way my students do.

I am allowed to select the measures that will be assessed for student growth

The evaluation doesn't really address music - it basically addresses my teaching ability.

I enjoy that I'm evaluated informally through every performance

i am not micro-managed by my administration

I am pretty much left to do my job, which is helpful to not have someone coming in who does not understand music/skills/techniques telling me what to do.

I appreciate the follow up conversation that takes place after the evaluation.

Administrators know nothing about what I teach

Our eval is a standard form for all teachers - no specifics to music; however general areas cross over in terms of content knowledge, classroom engagement, SEL, environment, etc.

An evaluator who understands and values music would have a positive impact. So would some attempt by the evaluator to understand and value my 50 years of music teaching experience.

High level of respect between myself and administrators. Allows me to develop my program and share goals with my administrative team.

Good teaching is good teaching, regardless of content knowledge on the part of the evaluator. If the evaluator recognizes good teaching practices, their lack of content knowledge should not matter.

They allow me to have input on how I measure student success.

Student formal & informal assessment; Contact with the community & parents;

I am allowed to set my own goals and have flexibility in the data which is used for the evaluation.

None

My Evaluator is very thorough and we have a pre-assessment meeting, an observation and then a follow-up post, observation meeting. My administrator works to ensure that these each happen in the same week so that the teacher is able to reflect immediately on their teaching.

My principals are very nice and ask me lots of questions to understand why I do certain things in my classroom since they don't understand music.

Relative feedback through an observation by a person who understands music education

Post-observation conferences provided helpful feedback regarding what the building administrator observed during my lesson and class time.

The building administrator (principal/vice-p.) are engaged through discussion and collaboration before and after the evaluation.

Informative meeting with my principal afterwards. Notice of one week before evaluation.

Advanced notice of evaluation, observation without distracting students' learning

The rubric used to collect data via direct observation evaluates what I do & how I do it, as opposed to checking off boxes.

Conversation with my Admin about the specifics of my class in general. My thoughts about the structures and my content specific issues.

Q21 List some factors, practices, or characteristics that have a negative impact on your levels of satisfaction with music teacher evaluation.

List some factors, practices, or characteristics that have a negative impact on your levels of satisfaction with music teacher evaluation.

A good portion of the language used in the evaluation process is not relevant to music education also my evaluators have no music experience (even when they were high school students) so they do not understand very well what I do in my classroom.

Occasionally, if the observation is unable to happen in school, some administrators use a concert as a method of evaluation.

Lack of knowledge of subject area by administrator. Little regard for teacher effectiveness.

The building admin don't really know what I'm doing as far as content.

the scores are limited based on what is expected. The evaluation is usually completed last minute

Sometimes, we are forced to "shoehorn" current educational trends in evaluation into the evaluation process, even if they have nothing to do with us.

In general school admin knows nothing about music.

Not having an admin with a knowledge of music give me feedback that's relevant to what I'm doing. My class looks very different than other core classes, and it would be helpful to get information about how to improve the aspects of my teaching related to music education, not just education general practices.

not sure how to answer

Overall, it seems like my classes are never really observed and they keep changing the way observations are handled.

finding time during planning to prepare data

Amount of evaluations

Lack of legitimate feedback, evaluations only at performances

I don't feel that there are any negative impact on my levels of satisfaction.

Sometimes they consider a concert an observation, so I don't know they are there, and I am unaware I being observed.

NA

The administrator who evaluates me often has no music experience or training.

They do not know my content

Coming in during a short portion and trying to get an overall picture of the class from 10-15 minutes.

NONE

My school system does not have a fine arts supervisor so I am observed and evaluated by someone who does not understand my content. This is frustrating because while they can comment and give feedback on posted learning intentions and standards and observe if students are engaged, they cannot advise me on ways to teach my content better. I am rarely notified ahead of time about observations, which is obviously a bit nerve-wracking and annoying because there are lots of time I am actively teaching, but lots of other times where we are just running through pieces and that is not the best representation of my teaching and lesson preparation. Additionally at my current school I have never had an in-personal debrief after an observation or received feedback on my goals or data. My school seems to operate on a "we will only meet with you if there is a problem" basis, I would love to actually have a conversation with my admin rather than being brushed aside so he can focus on a tested class.

Performance objectives (I.e. adjudicated assessments) cannot be used for evaluation purposes.

It doesn't actually address serious music education.

We are evaluated the same no matter subject area. It doesn't always apply or fit what we do

Administration not always attending my performances to see growth or successes.

It is not specific to music teaching or rehearsal practices.

Administrators know nothing about what I teach

Person observing does not have adequate training or background in my content area or how it may fit into the one size fits all evaluation form. It also doesn't take into consideration national music/arts standards

I do not have anyone in this building who has any idea what I am doing or what I am supposed to be doing other than myself.

Music teaching cannot be meaningfully evaluated by by someone outside the music profession. Specific criticisms of my situation include rushed and unsympathetic evaluation meetings, cancelled evaluation meetings that are never re-scheduled, refusals of data collection proposals for arbitrary reasons caused by the evaluators complete inability to understand them, unclear evaluation requirements, refusal to explain or elucidate questions, expecting busy colleagues to answer questions the evaluator cannot.

Having to fit the mold of non-music classrooms. Taking the time to creatively fit or explain how what I do fits into the evaluation form.

Without inclusion of student or teacher growth measures within my content area, there is little incentive on my part to learn, grow, change, or take risk.

Feedback from a music professional standpoint is limited.

Lack of evaluations don't allow principal to see the many levels on which I teach every day. He really needs to come into the classroom for an entire day because my classes are on different levels.

I don't think any administrator actually understands what we do...therefore it's easy to change/manipulate the data as needed.

My evaluating administrator has no idea what he is observing and has told that to me.

The evaluator is not a musician. This causes the evaluation to be an overview that is standard for every teacher...but there are some specific standards that are difficult to "meet" as a music teacher because they are not as applicable to the subject area.

Having principals that don't know anything about music so they truly can't effectively evaluate me.

lack of clear goals provided for me to reach for following an observation, lack of follow-ups

Some years, there was a high focus on data and student achievement, other years there was not. It would be more helpful for me to make the goals for my students, or for the students to make their own goals, and then to evaluate my success on achieving those goals.

When there is no real discussion about the evaluation.

I didn't see the rubric beforehand of how I would be evaluated.

The evaluator has limited or no knowledge of the content being taught. Mostly evaluated on classroom management and "flow" of the class period, so there is no constructive feedback on how to improve in content instruction.

Administrators rarely know very much about music & music education. They aren't able to offer much aside from praise.

The expectation for my classroom to be similar to a science or a reading class room.

Q19 In your experience, what are some characteristics found in a “good” evaluator?

In your experience, what are some characteristics found in a “good” evaluator?

They focus on the details of how I teach.

Knowledge of subject area, supportive, offers suggestions for improvement

A great leader, non-judgmental

Someone who asks questions and completed multiple observations

Someone who understands that we are "different:" our classes operated differently, instructional delivery will look and sound different, evaluation and feedback will be different. Someone who can see that and embrace it.

understanding what I do and or am supposed to do on a daily basis

Evaluators who take time to understand you, your subject, and what you're doing in order to give you the most appropriate and detailed feedback. I don't need you to necessarily be an expert, but at least ask questions/provide time before an evaluation to be educated on what's going to happen so you're not confused.

sticks to the facts of what happens in the classroom

Listens, supports, gives feedback with a level of empathy.

Someone who understands the art of teaching a music class, instead of a tested subject.

Understanding, adaptive, learning, gives the benefit of the doubt, looks for details, on your side/not "out to get you"

Understanding of content that they are evaluating.

A "good" evaluator is someone who is not just checking boxes. It is someone who see the overall picture of student and teacher growth and is there to support and also to challenge on ideas. A good evaluator works with you to grow and improve.

Ability to understand how a performing arts class can veer off a planned lesson because of the particular learning that is happening on that day.

One who understands good teaching practices are the same in every area.

Experience in the subject area being evaluated, time invested in becoming familiar with the teacher and thr program

They ask questions after the lesson

Objectiveness, understanding, classroom experience.

Understands that, just like students, there is not a one size fits all for teachers.

prior notification of observations, interactions with students during observations, in-person debrief after observstions, not just an email with google form repsonses. A good evaluator seeks to understand the content being taught and how data is collected.

Fair and consistent evaluation

Series of evaluations to assess actual growth in the subject matter.

They give constructive feedback that will make me a better educator all around

someone who takes the time to learn about your subject and what goes into it to be successful, not what things you are trying to cram into it to fit some kind of model.

Deferring to our expertise, having an understanding that our claims are run differently than core classes

Someone who has been in classrooms recently, checks in at least once a month with you.

Competent teacher, understands teaching in fine arts context.

Content knowledge

Open mind, experience in the subject matter, constructive feedback, being honest in the eval process and not just checking boxes to appease higher admin requirements.

Someone who knows how to tell me if I'm doing a good job or not, instead of someone who can only tell me whether or not my students are seated and paying attention.

Relevant experience in Music Education; objectivity; freedom from hidden (or overt) political agendas.

Open to ideas, good communication skills, willing to learn

Good teaching is good teaching, regardless of content. "Good" evaluators recognize engagement looks different in every classroom.

Willing to listen, open mind

Knowledge, understanding, and flexible.

A good evaluator has experience in the area he/she is evaluating or willing to admit his/her deficits in the post observation

Knowledge of the content area, understanding and compassion

They should have experience in the content knowledge of music education

A "good" evaluator knows the content area and relies on the the standards to formerly assess good teaching practices. A "good" evaluator understands that each classroom culture is unique and allows for a teacher to explain their methods or as my evaluator says "tell your story". An evaluator will be more effective if they have an understanding of the content area.

Someone who is observant, patient, willing to take the time and ask questions when they don't understand something.

The observer stays for the whole lesson/period. Evaluator has awareness of the history of the program

An evaluator who provided helpful feedback and specific insight to the structure and effectiveness of my lesson.

1. The evaluator asks questions about things that were uncertain; 2. The evaluator gives not only constructive feedback, but also highlights the positives as well; 3. The evaluator is not sexist.

Knowledge of subject matter, objectivity, classroom experience

Encouraging, positive, helps to improve my teaching

Someone who actually knows your content area and has experience teaching in that content area.

The person has had a wide variety of experience in many areas, including the arts.

self-awareness (acknowledging what they don't know about the subject being taught), empathy, ability to make constructive comments

Willingness to be ask questions about what they should see, hear, experience in a performance based class.

Q22 What advice would you give to new music teachers to your campus regarding music teacher evaluations?

What advice would you give to new music teachers to your campus regarding music teacher evaluations?

Understand that they do not understand what you do and will be to provide much advice or insight except in general classroom management and teacher practice. You may also have to spend time afterwards explaining things and proving that you do follow and do everything that is asked of you as a teacher and you meet all of the standards even if they cannot easily see it in class.

Be open-minded. Receive and embrace evaluations with idea that you will only grow.

Strive to made administrators with little or no music background with knowledge of what to look for.

Just do your job well and they will leave you alone.

Find out when and how you will be evaluated. Talk to a mentor about how to show what is expected. Ask questions!

Our evaluator is fair and is looking to see student engagement and success. He understands that we aren't math

Trust your own skills. teach well and admin can figure it out.

You're not going to be evaluated by someone who is an expert in music, so focus more heavily on standard practices that exist in good classrooms since that's what the evaluator will be looking most for.

Be organized. Document what you do.

Be comfortable with who you are as a teacher and don't "try" to do new things or things that are not you in observed lessons.

Do not stress about your evaluation. Take the first one and learn what your evaluator likes and do what they like for your next evaluation.

Plan a wise SMART goal that you KNOW you can achieve.

Find a teacher in your content to evaluate you if your assigned evaluator does not have experience.

Don't worry about it because it doesn't really matter

If you work every day on being the best teacher you can be, providing a safe and nurturing environment for your student, teacher evaluations become part of what you already do.

I would say to keep doing your plan, don't change based on an administrator walking in the room. If you had planned something crazy... do something crazy!

Talk to other teachers, regardless of content area, concerning areas that need improvement.

Be careful to set simple, measurable goals for student achievement.

Don't worry about them

Go about your teaching as normal. Don't do anything special to make your music class fit into the cookie cutter of a normal class.

Relax and be yourself.

If you want real advice on how to improve as a music teacher, ask a music teacher to observe you. Other than that, the only the admin seems to care about is are your standards, learning intentions, and success criteria posted, so try to at least have an agenda on the board and make sure you state goals for each rehearsal chunk and then remember to circle back at the end to check in with student to see how they feel about the concepts taught.

Do not use music terminology when writing the goals. Ensure they are in administrative friendly language

Know your content area and be faithful to appropriately teach the discipline of music.

Don't change how you teach when someone is in the room.

Do your job and you will be OK

The evaluation is about your teaching ability, not music teaching ability per se. there is great feedback and follow up with the setting of goals.

They are meaningless

Be sure to talk with mentor teachers in your school to understand the 'look for's' each school is different in how admin perceive your role

take them with a grain of salt, just make sure your classroom management is solid

Expect nothing of benefit; meet all deadlines; provide data that allows the evaluator to check the necessary boxes.

Don't fight the evaluation process, find how you fit into it.

Do not tailor your teaching to the evaluation. Engage in good teaching practices, and the evaluation will take care of itself.

Seek out other local music professionals for assistance and feedback.

Take them with a grain of salt. Have a music teacher evaluate you.

Talk to others in the same county so you can learn the process for music!

Ignore them as they are not helpful in any way

"Provide evidence of your student growth and examples of the ways you are implementing standards." "Understand that the evaluator has no idea what musical topics you are talking about or teaching, so as long as you are being an effective teacher and can explain your reasonings and methods he will be happy."

I would tell new music teachers to reach out to other music teachers when they need help. Also to know that evaluators don't understand what we do in our classrooms so their advice or criticism doesn't always apply to us.

Don't stress about them, just pick something that you are familiar with and like teaching and make sure the students are having fun.

Just do your very best every day and challenge your students to be the best they can be. Evaluations do not measure the sum of a person, but they can sometimes provide helpful insight for improvement.

Be yourself and teach with intentionality, as if you were being observed everyday!

Be prepared for dropins

Ask to see the evaluation criteria beforehand.

Have enough planned to last the entire class period, transition between activities quickly, sing/play more than you talk/give instruction

Take them with a grain of salt because usually whoever is evaluating you has no idea what you do every day.

Remember that YOU are the expert in the room on YOUR lesson. Don't sweat it if something does not go the way you wanted it to during an observation. Use the whole evaluative process to help streamline and organize your approach to lesson planning and instruction.

Be proactive as to what the admin should see. Be willing to be on the offense and send success and good performances to the admin.

Q23 What advice would you give to evaluators to your campus regarding music teacher evaluation?

What advice would you give to evaluators to your campus regarding music teacher evaluation?

Band content has some differences from the core classes that you are used to evaluating. Some of the things you suggest or think are missing either do not work well in our context or are not relevant to our subject matter.

Focus on the small things that occur during the rehearsal process. The end result is just a small component of teach music.

Look beyond process and look at product.

Let the fine arts supervisor do it instead of you.

Have the evaluation of music educators completed by someone trained in music education

Signs of success include student engagement and improvement. Look for direct and timely feedback that influences the flow of the rehearsal as needed.

Learn about the subject matter.

Ask more questions. Take more time to learn the complexities of what we're doing in order to be more successful in helping us grow as teachers. Just because I've received evaluations of "highly effective" each year I've been on eval cycle, I don't think that my growth and development as a teacher can be attributed to my evaluators. It has come much more from myself furthering my own education and seeking out opportunities to be educated with and by other peers in my field.

not sure

Make sure you know the national standards and criteria for music - it is a different animal than other subjects.

Give your teachers the benefit of the doubt. Look for what they are doing well and highlight that before seeking areas of improvement.

Try to find teachers that have a strong reputation, and go and watch them.

Take initiative to learn about what we do

Continue to support your teachers with good productive conversations on details of what makes outstanding educators.

Be flexible and try not to put our "square peg" lessons into the "round hole" of the rigid evaluation sheets.

NA

Take time to become familiar with the music program and the teachers

Ask more questions

The music classroom is different than a regular classroom.

Ask questions about things that you don't understand.

Please understand that music teachers cover many different standards in each so we don't always talk about every single one. Please interact with our students during observations, ask them what they are learning and how it relates to what they did yesterday and what they will do in 3 weeks from now. I know some admin try to measure teachers by the growth of the program, but not all school schedules are conducive to rapid growth. Instead talk to students about how they feel in class and how they feel about their teacher and their personal and group accomplishments in the program. This will give you a better idea of whether you have the right person in the job.

Performances need to be used as observations, as I am being observed by the entire community

It would be preferable for the evaluator to have some knowledge of the content area in order for evaluation of a music teacher to be helpful.

Ask questions and try to become knowledgeable about what is going on

Try to have a music supervisor or experienced music teacher conduct the evaluations.

Evaluators should understand content

consult with people who understand the discipline, ask questions for clarification, do not make assumptions

Please hire someone that can properly evaluate music & arts teachers, or at least provide adequate training for evaluations in the arts field.

Ask yourself, 'Why am I evaluating this teacher when I know nothing about Music and/or Music Education?'

Understand that we have the same goals and practices of "core teachers" but it looks very different.

Consider what good teaching practices look like in an ensemble setting. How do individual students make contributions to the larger whole?

Continue to be up to date in the current practices of the teaching profession.

Have a music teacher evaluate the teacher

Look for student engagement and attend performances!

Bring in someone with knowledge of the content to observe and evaluate

Learn what the musical terms mean. The concepts we teach are not difficult for an adult to pick up. Take a few minutes to look over our music standards and come into evaluations with the for-knowledge that allows for you to understand what I am doing and why.

Would it be possible to have another music teacher evaluate us instead of a person not competent in music education.

Be familiar with what the program history at least over the last few years. Ask ahead of time for specific areas that the teacher finds challenging so that the evaluator can try to offer insight in those specific areas.

It would be most helpful for a music educator or supervisor who is knowledgeable about my subject to evaluate my effectiveness as a music teacher.

When things are uncertain, ask the teacher in private after the observation.

Give advanced notice

If they are not knowledgeable in music, please ask questions afterwards.

Be willing to learn a little about the content being taught, meet with teacher before and after the observation.

Research to understand some of the content area and practices when observing music teachers.

If you don't have a background in music/music ed, defer to the expertise of the teacher being evaluated---turn any critical comments into an opportunity to collaborate with the teacher in question so that the teacher improves their performance and you grow in your understanding of music education.

If you don't know what you are looking at, or hearing, Ask.

APPENDIX D: LINEAR REGRESSION ANALYSIS ON TEACHER EVALUATION
PROCESS

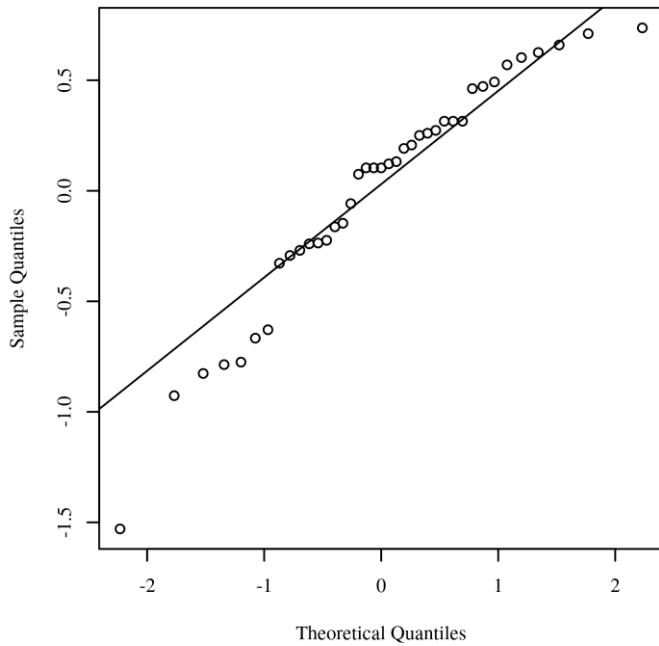
Linear Regression Analysis on Teacher Evaluation Process

A linear regression analysis was conducted to assess whether Q13.1, Q13.2, Q13.3, Q13.4, Q13.5, Q13.6, Q13.7, and Q13.8 significantly predicted Q13.9.

Assumptions

Normality. The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot (DeCarlo, 1997). For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Figure 4 presents a Q-Q scatterplot of the model residuals.

Figure D4. Q-Q scatterplot for normality of the residuals for the regression model



Multicollinearity. Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. High VIFs indicate increased effects of multicollinearity in the model. VIFs greater than 5 are cause for concern, whereas VIFs of 10 should be considered the maximum upper limit (Menard, 2009). All predictors in the regression model have VIFs less than 10. Table 30 presents the VIF for each predictor in the model.

Table D30. Variance Inflation Factors for Q13.1, Q13.2, Q13.3, Q13.4, Q13.5, Q13.6, Q13.7, and Q13.8

Variable	VIF
Q13.1	1.82
Q13.2	2.07
Q13.3	1.91
Q13.4	2.13
Q13.5	2.22
Q13.6	1.26
Q13.7	1.16
Q13.8	2.01

Results

The results of the linear regression model were significant, $F(8,30) = 20.14, p < .001, R^2 = .84$, indicating that approximately 84.30% of the variance in Q13.9 is explainable by Q13.1, Q13.2, Q13.3, Q13.4, Q13.5, Q13.6, Q13.7, and Q13.8. Q13.1 did not significantly predict Q13.9, $B = 0.34, t(30) = 1.82, p = .079$. Based on this sample, a one-unit increase in Q13.1 does not have a significant effect on Q13.9. Q13.2 did not significantly predict Q13.9, $B = 0.18, t(30) = 1.43, p = .162$. Based on this sample, a one-unit increase in Q13.2 does not have a significant effect on Q13.9. Q13.3 did not significantly predict Q13.9, $B = 0.02, t(30) = 0.19, p = .853$. Based on this sample, a one-unit increase in Q13.3 does not have a significant effect on Q13.9. Q13.4 significantly predicted Q13.9, $B = 0.29, t(30) = 2.52, p = .017$. This indicates that on

average, a one-unit increase of Q13.4 will increase the value of Q13.9 by 0.29 units. Q13.5 did not significantly predict Q13.9, $B = -0.30$, $t(30) = -1.85$, $p = .075$. Based on this sample, a one-unit increase in Q13.5 does not have a significant effect on Q13.9. Q13.6 did not significantly predict Q13.9, $B = -0.06$, $t(30) = -0.87$, $p = .389$. Based on this sample, a one-unit increase in Q13.6 does not have a significant effect on Q13.9. Q13.7 did not significantly predict Q13.9, $B = 0.05$, $t(30) = 0.73$, $p = .473$. Based on this sample, a one-unit increase in Q13.7 does not have a significant effect on Q13.9. Q13.8 significantly predicted Q13.9, $B = 0.59$, $t(30) = 6.06$, $p < .001$. This indicates that on average, a one-unit increase of Q13.8 will increase the value of Q13.9 by 0.59 units. Table 31 summarizes the results of the regression model.

Table D31. Results for Linear Regression with Q13.1, Q13.2, Q13.3, Q13.4, Q13.5, Q13.6, Q13.7, and Q13.8 predicting Q13.9

Variable	<i>B</i>	<i>SE</i>	95.00% CI	β	<i>t</i>	<i>p</i>
(Intercept)	-0.65	0.80	[-2.29, 0.99]	0.00	-0.81	.425
Q13.1	0.34	0.19	[-0.04, 0.73]	0.18	1.82	.079
Q13.2	0.18	0.13	[-0.08, 0.44]	0.15	1.43	.162
Q13.3	0.02	0.09	[-0.17, 0.21]	0.02	0.19	.853
Q13.4	0.29	0.12	[0.06, 0.53]	0.27	2.52	.017
Q13.5	-0.30	0.16	[-0.62, 0.03]	-0.20	-1.85	.075
Q13.6	-0.06	0.07	[-0.21, 0.09]	-0.07	-0.87	.389
Q13.7	0.05	0.07	[-0.10, 0.20]	0.06	0.73	.473
Q13.8	0.59	0.10	[0.39, 0.78]	0.62	6.06	< .001

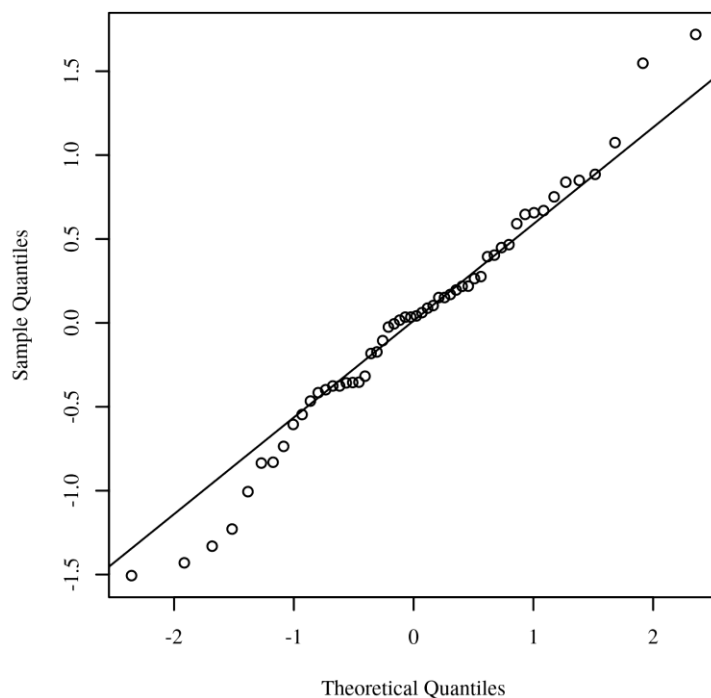
Linear Regression Analysis on Teacher Evaluation Personnel

A linear regression analysis was conducted to assess whether Q15.1, Q15.2, Q15.3, Q15.4, and Q15.5 significantly predicted Q15.6.

Assumptions

Normality. The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot (DeCarlo, 1997). For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Figure 5 presents a Q-Q scatterplot of the model residuals.

Figure D5. Q-Q scatterplot for normality of the residuals for the regression model.



Multicollinearity. Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. High VIFs indicate increased effects of multicollinearity in the model. VIFs greater than 5 are cause for concern, whereas VIFs of 10

should be considered the maximum upper limit (Menard, 2009). All predictors in the regression model have VIFs less than 10. Table 32 presents the VIF for each predictor in the model.

Table D32. Variance Inflation Factors for Q15.1, Q15.2, Q15.3, Q15.4, and Q15.5

Variable	VIF
Q15.1	1.40
Q15.2	1.97
Q15.3	2.17
Q15.4	1.54
Q15.5	2.15

Results

The results of the linear regression model were significant, $F(5,48) = 23.66, p < .001, R^2 = .71$, indicating that approximately 71.14% of the variance in Q15.6 is explainable by Q15.1, Q15.2, Q15.3, Q15.4, and Q15.5. Q15.1 did not significantly predict Q15.6, $B = -0.33, t(48) = -1.20, p = .235$. Based on this sample, a one-unit increase in Q15.1 does not have a significant effect on Q15.6. Q15.2 significantly predicted Q15.6, $B = 0.55, t(48) = 4.43, p < .001$. This indicates that on average, a one-unit increase of Q15.2 will increase the value of Q15.6 by 0.55 units. Q15.3 did not significantly predict Q15.6, $B = 0.20, t(48) = 1.99, p = .053$. Based on this sample, a one-unit increase in Q15.3 does not have a significant effect on Q15.6. Q15.4 did not significantly predict Q15.6, $B = 0.04, t(48) = 0.43, p = .666$. Based on this sample, a one-unit increase in Q15.4 does not have a significant effect on Q15.6. Q15.5 significantly predicted Q15.6, $B = 0.27, t(48) = 2.62, p = .012$. This indicates that on average, a one-unit increase of Q15.5 will increase the value of Q15.6 by 0.27 units. Table 33 summarizes the results of the regression model.

Table 33. Results for Linear Regression with Q15.1, Q15.2, Q15.3, Q15.4, and Q15.5 predicting Q15.6

Variable	<i>B</i>	<i>SE</i>	95.00% CI	β	<i>t</i>	<i>p</i>
(Intercept)	1.28	1.17	[-1.07, 3.62]	0.00	1.09	.279
Q15.1	-0.33	0.28	[-0.88, 0.22]	-0.11	-1.20	.235
Q15.2	0.55	0.13	[0.30, 0.81]	0.48	4.43	< .001
Q15.3	0.20	0.10	[-0.002, 0.41]	0.23	1.99	.053
Q15.4	0.04	0.10	[-0.16, 0.25]	0.04	0.43	.666
Q15.5	0.27	0.10	[0.06, 0.48]	0.30	2.62	.012

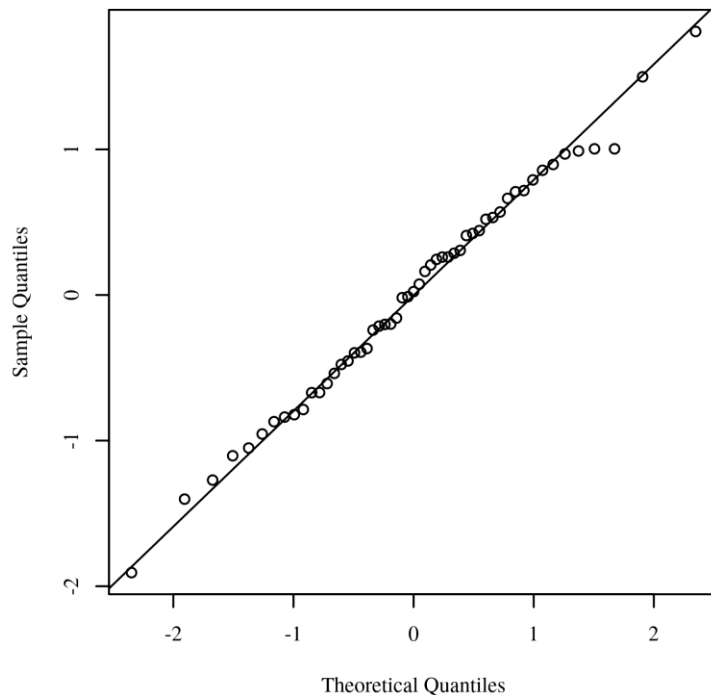
Linear Regression Analysis on Teacher Evaluation Product

A linear regression analysis was conducted to assess whether Q18.1, Q18.2, Q18.3, Q18.4, Q18.5, and Q18.6 significantly predicted Q18.7.

Assumptions

Normality. The assumption of normality was assessed by plotting the quantiles of the model residuals against the quantiles of a Chi-square distribution, also called a Q-Q scatterplot (DeCarlo, 1997). For the assumption of normality to be met, the quantiles of the residuals must not strongly deviate from the theoretical quantiles. Strong deviations could indicate that the parameter estimates are unreliable. Figure 6 presents a Q-Q scatterplot of the model residuals.

Figure D6. Q-Q scatterplot for normality of the residuals for the regression model.



Multicollinearity. Variance Inflation Factors (VIFs) were calculated to detect the presence of multicollinearity between predictors. High VIFs indicate increased effects of multicollinearity in the model. VIFs greater than 5 are cause for concern, whereas VIFs of 10

should be considered the maximum upper limit (Menard, 2009). All predictors in the regression model have VIFs less than 10. Table 34 presents the VIF for each predictor in the model.

Table D34. Variance Inflation Factors for Q18.1, Q18.2, Q18.3, Q18.4, Q18.5, and Q18.6

Variable	VIF
Q18.1	1.60
Q18.2	1.80
Q18.3	2.08
Q18.4	1.50
Q18.5	3.08
Q18.6	2.87

Results

The results of the linear regression model were significant, $F(6,46) = 13.07, p < .001, R^2 = .63$, indicating that approximately 63.04% of the variance in Q18.7 is explainable by Q18.1, Q18.2, Q18.3, Q18.4, Q18.5, and Q18.6. Q18.1 significantly predicted Q18.7, $B = 0.53, t(46) = 3.91, p < .001$. This indicates that on average, a one-unit increase of Q18.1 will increase the value of Q18.7 by 0.53 units. Q18.2 did not significantly predict Q18.7, $B = 0.26, t(46) = 1.90, p = .063$. Based on this sample, a one-unit increase in Q18.2 does not have a significant effect on Q18.7. Q18.3 did not significantly predict Q18.7, $B = -0.13, t(46) = -0.96, p = .342$. Based on this sample, a one-unit increase in Q18.3 does not have a significant effect on Q18.7. Q18.4 did not significantly predict Q18.7, $B = -0.10, t(46) = -1.09, p = .282$. Based on this sample, a one-unit increase in Q18.4 does not have a significant effect on Q18.7. Q18.5 did not significantly predict Q18.7, $B = 0.03, t(46) = 0.22, p = .829$. Based on this sample, a one-unit increase in Q18.5 does not have a significant effect on Q18.7. Q18.6 significantly predicted Q18.7, $B = 0.39, t(46) = 2.69, p = .010$. This indicates that on average, a one-unit increase of Q18.6 will increase the value of Q18.7 by 0.39 units. Table 35 summarizes the results of the regression model.

Table D35. Results for Linear Regression with Q18.1, Q18.2, Q18.3, Q18.4, Q18.5, and Q18.6 predicting Q18.7

Variable	<i>B</i>	<i>SE</i>	95.00% CI	β	<i>t</i>	<i>p</i>
(Intercept)	-0.09	0.59	[-1.28, 1.10]	0.00	-0.15	.881
Q18.1	0.53	0.13	[0.26, 0.80]	0.44	3.91	< .001
Q18.2	0.26	0.14	[-0.02, 0.54]	0.23	1.90	.063
Q18.3	-0.13	0.13	[-0.39, 0.14]	-0.12	-0.96	.342
Q18.4	-0.10	0.10	[-0.30, 0.09]	-0.12	-1.09	.282
Q18.5	0.03	0.14	[-0.26, 0.32]	0.03	0.22	.829
Q18.6	0.39	0.14	[0.10, 0.68]	0.41	2.69	.010